MULTIPLE SPANNWEITENTESTS

DISSERTATION

zur Erlangung des Grades
eines Doktors der Naturwissenschaften

Dem Fachbereich IV der Universität Trier
vorgelegt von

HELMUT FINNER

aus Düren

Trier 1988

UNIV.
LEEDS
SOC.

MULTIPLE SPANNWEITENTESTS

DISSERTATION

zur Erlangung des Grades
eines Doktors der Naturwissenschaften

Dem Fachbereich IV der Universität Trier
vorgelegt von

HELMUT FINNER

aus Düren

Trier 1988

Gutachter: Prof. Dr. E. Sonnemann, Universität Trier

Prof. Dr. H. Riedwyl, Universität Bern

Tag der mündlichen Prüfung: 05. Juli 1988

Inhalt

0. Einleitung und Übersicht 1

1. Multiple Tests 4

1.1. Grundbegriffe der Theorie der multiplen Tests 4

1.2. Der Abschlußtest und die Bedeutung der Kohärenz 7

1.3. Zwei Kriterien zur Wahl von multiplen Tests 12

2. Eigenschaften der Verteilungsfunktion der Spannweite 15

2.1. Die Verteilung der Spannweite 15

2.2. Log–Konkavität und Log–Konvexität der Spannweite 18

2.3. Eine grundlegende Ungleichung für die Verteilungsfunktion der Spannweite 22

3. Klassische multiple Spannweitentests 30

3.1. Modell, Zielhypothesen und ein klassischer Simultantest 33

3.2. Multiple Spannweitentests auf Basis der Homogenitätshypothesen 35

3.3. Das Problem der Monotonie der kritischen Werte 41

3.4. Eine universelle Wahl der adjustierten Signifikanzniveaus 44

3.5. Minimax– und Maximax–Prozeduren 48

4. Abgeschlossene multiple Spannweitentests 54

4.1. Wahl der lokalen Tests 54

4.2. Ordnungseigenschaften der lokalen Tests 60

4.3. Ein Algorithmus 73

4.4. Der abgeschlossene Newman–Keuls Spannweitentest im ANOVA–Modell mit gleichen Stichprobenumfängen 75

5. Abschlußbemerkungen 83

Literaturverzeichnis 86

Symbole und Bezeichnungen 90

0. Einleitung und Übersicht

Sind in einem statistischen Experiment mehrere Hypothesen zu prüfen, so spricht man von einem multiplen Testproblem. Zu den ältesten und in der Literatur meist betrachteten Problemen dieser Art gehört der Vergleich von k Mittelwerten im ANOVA–Modell. Ein typisches Beispiel ist der folgende von Keuls (1952) analysierte landwirtschaftliche Sortenversuch. Es sollten die Erträge von 13 verschiedenen Weißkohlsorten miteinander verglichen werden. Dazu wurden 3 Felder in jeweils 13 Parzellen aufgeteilt. Jeder Parzelle wurde zufällig eine der 13 Sorten so zugeteilt, daß jede Sorte auf jedem Feld genau einmal ausgebracht werden konnte. Es ergaben sich folgende mittlere Erträge pro Kopf für die verschiedenen Sorten (der Größe nach geordnet in 10 g):

Sorte	1	2	3	4	5	6	7	8	9	10	11	12	13
$\bar{x}_{i.}$	97.7	100.7	111.3	120.7	124.3	128.7	129.0	131.0	132.0	141.7	150.7	152.7	176.0

Während die klassische Varianzanalyse nur die Frage beantwortet, ob es überhaupt Unterschiede zwischen den Sorten gibt, will man im allgemeinen auch wissen, welche Sorten sich voneinander unterscheiden. Bezeichnet μ_i den theoretischen mittleren Ertrag der i–ten Sorte, so bedeutet diese Fragestellung in der Sprache der statistischen Testtheorie, daß die Hypothesen

$$(0.1) \qquad H_0^{ij} : \mu_i = \mu_j \text{ gegen } H_1^{ij} : \mu_i \neq \mu_j \,, \; 1 \leq i < j \leq 13$$

mit geeigneten statistischen Tests zu prüfen sind.

Wie beim Prüfen einer Hypothese stellt sich auch hier die Frage nach einer adäquaten Kontrolle der möglichen Fehlentscheidungen, insbesondere der Kontrolle der Fehler I. Art. Würde man im obigen Beispiel jede der $\binom{13}{2} = 78$ Nullhypothesen mit einem

t–Test zum Niveau α prüfen (z. B. $\alpha = 0.05$), so ist unmittelbar klar, daß die Wahrscheinlichkeit, irgendeine wahre Nullhypothese abzulehnen, wesentlich größer als α wird. Dieser Effekt wird durch die Einführung eines multiplen Niveaus α und eine entsprechende Wahl der Tests verhindert. Dabei sind die Tests so zu wählen, daß unabhängig davon, welche der Hypothesen wahr sind, die Wahrscheinlichkeit für die Ablehnung irgendeiner wahren Nullhypothese durch das vorgegebene α beschränkt wird.

Ein sehr allgemein anwendbares Konzept zur Konstruktion von Tests zum multiplen Niveau α wurde erstmals von Marcus, Peritz, Gabriel (1976) formuliert und führt zu den sogenannten sequentiell verwerfenden Testprozeduren bzw. zu den Abschlußtests. Ansätze dieses Konzepts sind auch schon in Einot, Gabriel (1975) und Naik (1975) zu finden. Obwohl sich mit Hilfe des aus der Arbeit von Marcus et al. abgeleiteten Abschlußprinzips (vgl. Sonnemann (1982)) viele klassische Tests verbessern lassen, scheitert eine Anwendung häufig an der Größe der resultierenden Testprobleme. So sind im Fall der Paarvergleiche bei einer konsequenten Anwendung des Abschlußprinzips bei 13 Mittelwerten formal 27 644 436 Hypothesen, bei 20 Mittelwerten mehr als $51 \cdot 10^{12}$ Hypothesen zu prüfen. Aus diesem Grund gibt man sich häufig mit suboptimalen Lösungen zufrieden.

Ziel der vorliegenden Arbeit ist es, speziell für den Vergleich von k Mittelwerten oder Lokationsparametern mit Hilfe multipler Spannweitentests vor dem Hintergrund des Abschlußprinzips Lösungen zu entwickeln. Zu diesem Zweck wurde die Arbeit in 4 mehr oder weniger in sich abgeschlossene Kapitel unterteilt.

Im Kapitel 1 werden zunächst die Grundbegriffe der Theorie der multiplen Tests eingeführt. Anschließend wird anhand einiger Sätze die Bedeutung des Abschlußprinzips und der Kohärenz aufgezeigt, bevor zwei Kriterien zur Wahl von multiplen Tests vorgestellt werden, die in der Literatur bisher nicht betrachtet wurden.

Das Kapitel 2 enthält einige Resultate für die Verteilungsfunktion der Spannweite. Insbesondere wird eine für beliebige Verteilungen gültige Ungleichung bewiesen, die beim Nachweis der Monotonie der kritischen Werte eines klassischen multiplen

Spannweitentests eine entscheidende Rolle spielt. Daneben wird ein Zusammenhang zur Log–Konvexität bzw. Log–Konkavität der Verteilungsfunktion der Spannweite hergeleitet.

Im Kapitel 3 beschäftigen wir uns zunächst mit klassischen multiplen Spannweitentests auf Basis der sogenannten Homogenitätshypothesen, die dem Newman–Keuls Test (Newman (1939), Keuls (1952)) nachempfunden sind, jedoch im Gegensatz zu diesem das multiple Niveau α einhalten. Insbesondere wird in diesem Teil der Arbeit ausgiebig auf die Forderung der Monotonie der kritischen Werte (Lehmann, Shaffer (1977)) und die "optimale" Wahl der adjustierten Signifikanzniveaus eingegangen (vgl. Lehmann, Shaffer (1979), aber auch Lehmann (1986), pp. 380–388). Dabei werden einige neue Ansätze diskutiert. Die Bedeutung der in diesem Kapitel vorgestellten Testverfahren, die sich alle mit Hilfe des Abschlußprinzips verbessern lassen, liegt in ihrer einfachen Anwendbarkeit.

Im 4. und letzten Kapitel wird schließlich ein Abschluß des Newman–Keuls Tests vorgestellt, der auf einer konsequenten Anwendung des Abschlußprinzips beruht. Im Gegensatz zu dem von Begun, Gabriel (1981) vorgestellten Abschluß des Newman–Keuls Tests im ANOVA–Modell werden für die Partitionshypothesen Maxima von Spannweitenstatistiken verwendet. Durch Ausnutzen von Ordnungseigenschaften der Spannweitenstatistiken wird die Zahl der zu prüfenden Hypothesen so reduziert, daß sich Probleme mit bis zu 20 Lokationsparametern mit Rechnerhilfe bequem lösen lassen. Insbesondere schöpft dieser Test durch die Verwendung der exakten Verteilungen das vorgegebene multiple Niveau α im Gegensatz zu allen anderen bekannten Verfahren voll aus.

Mein besonderer Dank gilt Prof. Dr. E. Sonnemann, der mich nicht nur zu diesem Thema ermutigte, sondern auch durch seine ständige Diskussionsbereitschaft eine große Unterstützung war. Bedanken möchte ich mich auch bei Christa Willems für ihre Geduld und Sorgfalt beim Schreiben dieser Arbeit.

1. Multiple Tests

Wir geben zunächst eine allgemeine Einführung in die Grundbegriffe der Theorie der multiplen Tests. Dabei beschränken wir uns auf die für diese Arbeit wichtigen Begriffe. Den Bonferroni–Holm Test etwa (Holm (1977,1979a)) oder eine seiner Varianten (vgl. Hommel (1986,1988), Shaffer (1986), Simes (1986)) wird der Leser vergeblich suchen. In solchen Fällen verweisen wir auf die Literatur, insbesondere auch auf die Übersichtsarbeit von Sonnemann (1982).

Im Mittelpunkt unserer Betrachtungen stehen das multiple Niveau α, das Abschlußprinzip, der Abschlußtest und der Kohärenzbegriff. Anhand einiger Sätze in Abschnitt 1.2 wird der Zusammenhang dieser Begriffe verdeutlicht. Das wohl wichtigste Resultat dieses Abschnitts ist die bisher nicht registrierte Tatsache, daß die Menge der Abschlußtests wesentlich vollständig (im Sinne der Entscheidungstheorie) in der Menge der (nicht–randomisierten) Tests ist (vgl. auch Sonnemann, Finner (1988)). In Abschnitt 1.3 wird eine prinzipielle Schwierigkeit bei der Konstruktion von multiplen Tests mit dem Abschlußprinzip diskutiert, die leicht übersehen wird. Schließlich werden zwei neue Begriffe eingeführt, die als Entscheidungshilfe bei der Wahl von multiplen Tests dienen sollen.

1.1. Grundbegriffe der Theorie der multiplen Tests

Sei X eine Zufallsvariable mit unbekannter Verteilung aus $\{P_\vartheta : \vartheta \in \Theta\}$. Ist $\mathcal{X}$ der zu X gehörende Stichprobenraum und $\mathcal{B}$ eine σ–Algebra über $\mathcal{X}$, so nennt man das Tripel $(\mathcal{X}, \mathcal{B}, \{P_\vartheta : \vartheta \in \Theta\})$ einen statistischen Raum. Ein multiples Testproblem ist dann gegeben, wenn über $\{P_\vartheta : \vartheta \in \Theta\}$ mindestens zwei zu prüfende Nullhypothesen ausgezeichnet sind. Ist I mit $|I| \geq 2$ eine beliebige Indexmenge, so bezeichnet $\mathcal{H} := \{H_0^i : i \in I\}$ mit $\emptyset \neq H_0^i \subset \Theta$ und $H_0^i \neq H_0^j$ für $i \neq j$ eine Familie von Nullhypothesen. Die entsprechenden Alternativhypothesen sind durch $H_1^i := \Theta \setminus H_0^i$ gegeben.

Für die einzelnen Testprobleme H_0^i gegen H_1^i definiert man (nicht-randomisierte) Tests $\varphi_i : (\mathcal{X}, \mathcal{B}) \to (\{0,1\}, 2^{\{0,1\}})$, wobei $\varphi_i = 1$ bzw. $\varphi_i = 0$ als Entscheidung für bzw. gegen H_0^i interpretiert wird. Entsprechend nennt man die Menge

$$\{\varphi_i = 0\} := \{x \in \mathcal{X} : \varphi_i(x) = 0\} \quad \text{bzw.}$$
$$\{\varphi_i = 1\} := \{x \in \mathcal{X} : \varphi_i(x) = 1\}$$

Annahme- bzw. Ablehnungsbereich des Tests φ_i für H_0^i. Der aus den Einzeltests gebildete Test $\varphi := (\varphi_i : i \in I)$ ist ein multipler Test für $\mathcal{H}$. Um Meßbarkeitsprobleme zu vermeiden, wird im folgenden endliches I vorausgesetzt.

Wie in der Neyman-Pearson Testtheorie spielt die Kontrolle der möglichen Fehlerarten eine zentrale Rolle. In dieser Arbeit wird nur das sich immer mehr durchsetzende Konzept der simultanen Kontrolle aller Fehler I. Art benutzt, das sich in der folgenden Definition widerspiegelt:

Definition 1.1. $\varphi = (\varphi_i : i \in I)$ heißt Test zum multiplen Niveau α für $\mathcal{H} = \{H_0^i : i \in I\}$, wenn gilt:

$$(1.1) \qquad \forall\, \vartheta \in \Theta : P_\vartheta\Big(\bigcup_{i \in I(\vartheta)} \{\varphi_i = 1\}\Big) \le \alpha .$$

Dabei sei $\alpha \in [0,1]$ und $I(\vartheta) := \{i \in I : \vartheta \in H_0^i\}$ die Menge der Indizes der unter ϑ wahren Nullhypothesen. Für $I(\vartheta) = \emptyset$ wird die Konvention $\bigcup_{i \in \emptyset} A_i = \emptyset$ verwendet.

Ist $\varphi = (\varphi_i : i \in I)$ ein Test zum multiplen Niveau α, so wird die Wahrscheinlichkeit, auch nur eine wahre Nullhypothese abzulehnen, durch das vorgegebenen Niveau α beschränkt. Sonnemann (1982) zeigt, daß (1.1) äquivalent ist mit

UNIV. LEIPZIG BIBL.

(1.2) $$\forall\, \emptyset \neq J \subseteq I : \forall\, \vartheta \in H_0^J := \bigcap_{j\in J} H_0^j : P_\vartheta\Big(\bigcup_{j\in J}\{\varphi_j = 1\}\Big) \leq \alpha .$$

Die Bedingung (1.2) besagt insbesondere, daß jeder Test zum multiplen Niveau α für $\mathcal{H}$ immer auch einen Test $\bar{\varphi} := (\varphi_i : i \in \bar{I})$ für

$$\begin{aligned}\bar{\mathcal{H}} &:= \{H_0^i : i \in \bar{I}\}\\ &:= \{H_0 \subset \Theta : \exists\, J \subseteq I : H_0^J = H_0 \neq \emptyset\}\end{aligned}$$

liefert. Wegen $\mathcal{H} \subseteq \bar{\mathcal{H}}$ bedeutet es keine Einschränkung, sogleich von einer Hypothesenfamilie $\bar{\mathcal{H}} := \{H_0^i : i \in I\}$ mit $H_0^i \neq H_0^j$ für $i \neq j$ auszugehen, die bezüglich der Durchschnittsbildung abgeschlossen ist.

Definition 1.2. Ein multiples Testproblem $(\mathfrak{X}, \mathfrak{B}, \{P_\vartheta : \vartheta \in \Theta\})$, $\{H_0^i : i \in I\}$ bzw. eine Familie von Nullhypothesen $\{H_0^i : i \in I\}$ mit $\emptyset \neq H_0^i \subset \Theta$ und $H_0^i \neq H_0^j$ für $i \neq j$ heißt durchschnittsabgeschlossen (kurz: $\cap$–abg.), wenn gilt:

(1.3) $$\forall\, \emptyset \neq J \subseteq I : H_0^J := \bigcap_{j\in J} H_0^j = \emptyset \text{ oder } H_0^J \in \{H_0^i : i \in I\} .$$

Bemerkung 1.1. Ist $\bar{\mathcal{H}} = \{H_0^i : i \in I\}$ $\cap$–abg. und $\varphi = (\varphi_i : i \in I)$ ein Test zum multiplen Niveau α für $\bar{\mathcal{H}}$, so folgt für alle $\emptyset \neq J \subseteq I$, daß $\varphi_J = (\varphi_i : i \in J)$ ein Test zum multiplen Niveau α für $\{H_0^i : i \in J\}$ ist. Das heißt, jeder Test zum multiplen Niveau α für eine $\cap$–abg. Hypothesenfamilie liefert einen solchen für jede Teilfamilie $\mathcal{H} \subseteq \bar{\mathcal{H}}$.

Eine Hypothese $H_0^i \in \bar{\mathcal{H}}$ nennt man Elementarhypothese, falls es keine echte Obermenge von H_0^i in $\bar{\mathcal{H}}$ gibt. Elementarhypothesen sind also die maximalen Elemente bzgl. der durch die Mengeninklusion gegebenen Halbordnung. Die minimalen Elemente in dieser Ordnung heißen Globalhypothesen. Eine Hypothese H_0^i heißt Schnitthypothese, falls sie Durchschnitt ihrer echten Obermengen ist.

Die vielfältigen Möglichkeiten, mit multiplen Tests nicht nur zu falschen, sondern zu widersprüchlichen Entscheidungen zu kommen, stellen ein Problem für sich dar (vgl. z.B. Lehmann (1957a,1957b)). Ein solcher Widerspruch ergibt sich, wenn ein H_0^i angenommen, eine Obermenge $H_0^j \supset H_0^i$ aber abgelehnt wird. Ein anderer kann bei Schnitthypothesen auftreten, die abgelehnt werden, deren sämtliche Obermengen aber akzeptiert werden. Um diese beiden Typen – und das sind nicht alle möglichen – zu kontrollieren, wurden von Gabriel (1969) die Begriffe Kohärenz und Konsonanz eingeführt.

Definition 1.3. (Gabriel (1969)). Ein multipler Test $\varphi = (\varphi_i : i \in I)$ für $\mathcal{H} = \{H_0^i : i \in I\}$ heißt

(a) kohärent, wenn gilt:

$$\forall\, i,j \in I : H_0^j \supset H_0^i \Rightarrow \{\varphi_j = 1\} \subseteq \{\varphi_i = 1\}.$$

(b) konsonant, wenn gilt:

$$\forall\, i \in I : [\exists\, H_0^j \supset H_0^i \Rightarrow \{\varphi_i = 1\} \subseteq \bigcup_{\ell : H_0^\ell \supset H_0^i} \{\varphi_\ell = 1\}].$$

Die Bedeutung der Kohärenz ergibt sich aus dem folgenden Abschnitt. Die Forderung nach der Konsonanz geht im allgemeinen auf Kosten der Güte eines multiplen Tests und ist aus diesem Grund häufig nicht erstrebenswert. Eine ausführliche Diskussion dieser und verwandter Begriffe findet man in Sonnemann (1982).

1.2. Das Abschlußprinzip und die Bedeutung der Kohärenz

Ein universelles Konzept zur Konstruktion von Tests zum multiplen Niveau α wurde erstmals von Marcus, Peritz, Gabriel (1976) vorgestellt. Ideen dieses Konzepts finden sich allerdings auch schon in Naik (1975).

Definition 1.4. Ein Test $\varphi = (\varphi_i : i \in I)$ heißt Test zum lokalen Niveau α für $\mathcal{H} = \{H_0^i : i \in I\}$, falls alle seine Komponenten Tests zum Niveau α sind, d.h., falls gilt:

$$(1.3) \qquad \forall\, i \in I : \forall\, \vartheta \in H_0^i : P_\vartheta(\varphi_i = 1) \leq \alpha .$$

Satz 1.1. (Marcus, Peritz, Gabriel (1976)). Ist $\mathcal{H} = \{H_0^i : i \in I\}$ $\cap$-abg. und $\varphi = (\varphi_i : i \in I)$ ein Test zum lokalen Niveau α für $\mathcal{H}$, so ist der Test $\bar{\varphi} := (\bar{\varphi}_i : i \in I)$ mit

$$(1.4) \qquad \forall\, i \in I : \bar{\varphi}_i := \begin{cases} 1 , & \forall\, H_0^j \subseteq H_0^i : \varphi_j = 1 \\ 0 , & \text{sonst} \end{cases}$$

ein Test zum multiplen Niveau α für $\mathcal{H}$.

Der durch (1.4) definierte Test $\bar{\varphi}$ heißt der zu φ gehörende Abschlußtest. Das Konstruktionsverfahren heißt Abschlußprinzip.

Einen ersten Zusammenhang zwischen dem Abschlußtest und der Kohärenz eines Tests zeigt der folgende

Satz 1.2. (Vgl. Sonnemann (1982)). Sei $\varphi = (\varphi_i : i \in I)$ ein Test zum lokalen Niveau α für die $\cap$-abg. Hypothesenfamilie $\mathcal{H} = \{H_0^i : i \in I\}$. Dann gilt:

(a) Der Abschlußtest $\bar{\varphi}$ ist kohärent.

(b) Ist φ kohärent, so ist φ Test zum multiplen Niveau α und es gilt: $\varphi = \bar{\varphi}$.

Dieser Satz legt die Frage nahe, ob nicht jeder in einem gewissen Sinne "gute" Test zum multiplen Niveau α kohärent und damit auch Abschlußtest sein muß. Als Vergleichskriterium verwenden wir den komponentenweisen Vergleich der Ablehnungsbereiche.

Definition 1.5. Seien $\varphi = (\varphi_i : i \in I)$ und $\varphi' = (\varphi'_i : i \in I)$ multiple Tests für $\mathcal{H} = \{H_0^i : i \in I\}$.

φ heißt gleichmäßig nicht schlechter als $\varphi' :\Leftrightarrow$

$$(1.5) \qquad \forall\, i \in I : \{\varphi_i = 1\} \supseteq \{\varphi'_i = 1\} \quad \text{bzw.} \quad \varphi_i \geq \varphi'_i .$$

Aus (1.5) folgt insbesondere, daß φ komponentenweise nicht kleinere Güte besitzt als φ'.

Satz 1.3. (Sonnemann, Finner (1988)). Sei $\overline{\mathcal{H}} = \{H_0^i : i \in I\}$ $\cap$-abg. und $\varphi = (\varphi_i : i \in I)$ ein Test zum multiplen Niveau α für $\overline{\mathcal{H}}$. Dann existiert ein Test $\varphi' = (\varphi'_i : i \in I)$ mit den Eigenschaften

(a) φ' ist kohärenter Test zum multiplen Niveau α für $\overline{\mathcal{H}}$.

(b) $\varphi' = \overline{\varphi'}$.

(c) φ' ist gleichmäßig nicht schlechter als φ.

Beweis: Definiert man

$$\varphi'_i := \begin{cases} 1 , & \exists\, H_0^j \supseteq H_0^i : \varphi_j = 1 \\ 0 , & \text{sonst} \end{cases} ,$$

dann folgt (c) wegen $\{\varphi'_i = 1\} = \bigcup_{j : H_0^j \supseteq H_0^i} \{\varphi_j = 1\} \supseteq \{\varphi_i = 1\}$.

Ist $H_0^\ell \supset H_0^m$, so gilt

$$\{\varphi'_\ell = 1\} = \bigcup_{j : H_0^j \supseteq H_0^\ell} \{\varphi_j = 1\} \subsetneq \bigcup_{j : H_0^j \supseteq H_0^m} \{\varphi_j = 1\} = \{\varphi'_m = 1\} ,$$

also die Kohärenz von φ'.

Da φ ein Test zum multiplen Niveau α ist, folgt für $i \in I$ und

$J_i := \{j \in I : H_0^j \supseteq H_0^i\}$ mit (1.2)

$$\forall\, \vartheta \in H_0^i = H_0^{J_i} := \bigcap_{j \in J_i} H_0^j : P_\vartheta(\varphi_i' = 1) = P_\vartheta\Big(\bigcup_{j \in J_i} \{\varphi_j = 1\}\Big) \le \alpha.$$

Das heißt jedoch, φ' ist kohärenter Test zum lokalen Niveau α, also nach Satz 1.2(b) auch Test zum multiplen Niveau α mit $\varphi' = \overline{\varphi'}$. □

Bemerkung 1.2. Satz 1.3 zeigt, daß man im Fall einer ∩–abg. Hypothesenfamilie $\mathcal{H}$ zu jedem Test φ zum multiplen Niveau α stets einen gleichmäßig nicht schlechteren kohärenten Test φ' zum multiplen Niveau α finden kann, der mit seinem Abschlußtest übereinstimmt. Das heißt, die Menge der Abschlußtests bzw. der kohärenten Tests zum multiplen Niveau α ist wesentlich vollständig (im Sinne der Entscheidungstheorie) in der Menge aller (nicht–randomisierten) Tests zum multiplen Niveau α. Ist $\mathcal{H}$ nicht ∩–abg., so läßt sich ebenfalls zeigen, daß die Menge der kohärenten Tests zum multiplen Niveau α wesentlich vollständig in der Menge aller Tests zum multiplen Niveau α für $\mathcal{H}$ ist. □

Ist $\mathcal{H} = \{H_0^i : i \in I\}$ eine gegebene Hypothesenfamilie und $\varphi = (\varphi_i : i \in I)$ ein kohärenter Test für $\mathcal{H}$, so stellt sich die Frage, unter welchen zusätzlichen Bedingungen φ ein Test zum multiplen Niveau α ist. Sei dazu

$$\overline{\mathcal{H}} := \{H_0^i : i \in \overline{I}\}$$

die durch $\mathcal{H}$ erzeugte ∩–abg. Hypothesenfamilie und

$$(1.6)\quad J(i) := \begin{cases} \{j \in I : H_0^j \supseteq H_0^i \text{ und } [\exists\, k \in I : H_0^j \supseteq H_0^k \supseteq H_0^i \Rightarrow k \in \{i,j\}]\}, & i \in \overline{I} \setminus I \\ \{i\} & , \ i \in I \ . \end{cases}$$

d.h., für $i \in \bar{I} \setminus I$ ist J(i) die Menge der Indizes der kleinsten Hypothesen in $\mathcal{H}$, die H_0^i gerade noch umfassen. Dadurch wird erreicht, daß sich jede Hypothesen $H_0^i \in \bar{\mathcal{H}}$ darstellen läßt als

$$H_0^i = H_0^{J(i)} := \bigcap_{j \in J(i)} H_0^j .$$

Der Vorteil dieser etwas komplizierteren Schreibweise wird deutlich, wenn man die Bedingung (c) des folgenden Satzes mit der Definition des multiplen Niveau α bzw. der Bedingung (1.2) vergleicht.

Satz 1.4. Sei φ ein kohärenter Test für $\mathcal{H} = \{H_0^i : i \in I\}$ und $\bar{\mathcal{H}}$, I, $\bar{I}$ und J(i) wie oben definiert. Dann sind folgende Aussagen äquivalent:

(a) φ ist Test zum multiplen Niveau α für $\mathcal{H}$.

(b) Der durch

$$\forall\, i \in \bar{I} : \bar{\varphi}_i := \begin{cases} 1 , & \exists\, j \in J(i) : \varphi_j = 1 \\ 0 , & \text{sonst} \end{cases}$$

definierte kohärente Test $\bar{\varphi} := (\bar{\varphi}_i : i \in \bar{I})$ ist Test zum multiplen Niveau α für $\bar{\mathcal{H}}$.

(c) $$\forall\, i \in \bar{I} : \forall\, \vartheta \in H_0^{J(i)} : P_\vartheta\Big(\bigcup_{j \in J(i)} \{\varphi_j = 1\}\Big) \le \alpha .$$

Beweis: Die Kohärenz von $\bar{\varphi}$ folgt aus Definition von $\bar{\varphi}$, J(i) und der Kohärenz von φ. Ist (c) erfüllt, so ist $\bar{\varphi}$ Test zum lokalen Niveau α. Aus der Kohärenz von $\bar{\varphi}$ folgt mit Satz 1.2.(b), daß $\bar{\varphi}$ Test zum multiplen Niveau α für $\bar{\mathcal{H}}$ ist, also (b). Ist (b) erfüllt, so folgt (a) wegen $\bar{\varphi}_i = \varphi_i$ für $i \in I$. Aus (a) folgt insbesondere (c). □

1.3. Zwei Kriterien zur Wahl von multiplen Tests

Obwohl das Abschlußprinzip ein schlagkräftiges Instrument zur Konstruktion von Tests zum multiplen Niveau α liefert, ist bei seiner Anwendung einige Vorsicht geboten. Insbesondere sollte man vor Anwendung von Satz 1.1. genau festlegen, welche Hypothesen wirklich von Interesse sind. So nützt es zum Beispiel wenig, die Güte einzelner Tests für Hypothesen von untergeordnetem Interesse auf Kosten der tatsächlich interessierenden Hypothesen zu erhöhen. Besonders groß ist diese Gefahr, wenn man nur an den Elementarhypothesen der $\cap$–abg. Hypothesenfamilie interessiert ist. Der Grund liegt darin, daß eine Elementarhypothese H_0^i beim Abschlußtest höchstens dann abgelehnt wird, wenn alle $H_0^j \subset H_0^i$ schon abgelehnt worden sind. Das heißt aber, daß sich das Niveau eines Tests $\bar{\varphi}_i$ für H_0^i im Verhältnis zum Niveau–α–Test φ_i drastisch verringern kann. Optimal wäre natürlich, wenn alle lokalen Tests $\bar{\varphi}_i$ des Abschlußtests $\bar{\varphi}$ das vorgegebene Niveau α voll ausschöpften. Das ist jedoch nur selten möglich, da das multiple Niveau α in gewisser Weise auf die einzelnen Hypothesen aufgeteilt wird. Zwei neue Begriffsbildungen sollen helfen, eine Art "innere Konsistenz" von multiplen Tests und ein möglichst vollständiges Ausschöpfen des multiplen Niveaus sicherzustellen.

Definition 1.6. Sei $\varphi = (\varphi_i : i \in I)$ ein Test zum lokalen Niveau α für $\mathcal{H} = \{H_0^i : i \in I\}$, $\overline{\mathcal{H}} = \{H_0^i : i \in \overline{I}\}$ die durch $\mathcal{H}$ erzeugte $\cap$–abg. Hypothesenfamilie und $\bar{\varphi} = (\bar{\varphi}_i : i \in \overline{I})$ der durch (1.4) definierte Abschlußtest für $\overline{\mathcal{H}}$.

(a) $\bar{\varphi}$ heißt lokal konsonant, falls φ konsonant ist.

(b) $\bar{\varphi}$ (bzw. φ, falls φ Test zum multiplen Niveau α ist) heißt das multiple Niveau α voll ausschöpfend, falls

$$\forall\, i \in \overline{I} : \sup_{\vartheta \in H_0^i} P_\vartheta(\bar{\varphi}_i = 1) = \alpha$$

$$\left(\text{bzw. } \forall\, i \in \overline{I} : \sup_{\vartheta \in H_0^i} P_\vartheta\Big(\bigcup_{j \in J_i} \{\varphi_j = 1\}\Big) = \alpha \text{ mit } J_i := \{j \in I : H_0^j \supseteq H_0^i\}\right).$$

Teil (b) dieser Definition ist insofern problematisch, daß in dieser Arbeit nur nichtrandomisierte multiple Tests betrachtet werden und ein Ausschöpfen des Niveaus i.a. den Einsatz randomisierter Tests erfordert. Randomisierte multiple Tests aber führen zu einigen technischen Problemen (vgl. Sonnemann, Finner (1988)), die nicht Gegenstand dieser Arbeit sein sollen. Deshalb werden in einigen Abschnitten Lebesgue–stetige Verteilungen vorausgesetzt.

Definiert man die lokalen Tests φ_i von $\varphi = (\varphi_i : i \in I)$ auf der Basis von Überschreitungswahrscheinlichkeiten, so erfüllt der resultierende Abschlußtest in vielen Fällen sowohl (a) als auch (b) der Definition 1.6. Als eine Menge von Tests, die (a), jedoch nicht immer (b) erfüllen, werden sich die multiplen Spannweitentests für Lokationsfamilien erweisen. Das folgende Beispiel zeigt, daß der Abschlußtest bei Verwendung von χ^2–Tests i.a. nicht lokal konsonant ist.

Beispiel 1.1. Seien $X_i \sim N(\mu_i,1)$, $i = 1,2$ stochastisch unabhängig und $\alpha = 0.05$. Zu prüfen seien die Hypothesen $H_0^i : \mu_i = 0$, $i = 1,2$ und $H_0^{12} : \mu_1 = \mu_2 = 0$. Für H_0^i , $i = 1,2$, seien die Niveau–α–Tests

$$\varphi_i = \begin{cases} 1 \\ 0 \end{cases} , \quad |x_i| \begin{matrix} > \\ \leq \end{matrix} u_{\alpha/2} \quad , \quad i = 1,2 ,$$

und für H_0^{12} der Niveau–α–Test

$$\varphi_{12} = \begin{cases} 1 \\ 0 \end{cases} , \quad x_1^2 + x_2^2 \begin{matrix} > \\ \leq \end{matrix} \chi^2_{2,\alpha}$$

gegeben. Wählt man $x = (x_1,x_2) = (1.9,1.9)$, so ist zwar $x_1^2 + x_2^2 = 7.22 > \chi^2_{2,\alpha} = 5.99$, aber $|x_i| = 1.9 < u_{\alpha/2} = 1.96$. Also ist $\varphi = (\varphi_1, \varphi_2, \varphi_{12})$ nicht konsonant. Ist man nur an den Hypothesen H_0^i, $i = 1,2$, interessiert, so führt das Vorschalten des χ^2–Tests für H_0^{12} dazu, daß unter H_0^{12} die Wahrscheinlichkeit, eine der Hypothesen

H_0^i, i=1,2, abzulehnen, echt kleiner als α ist. Andererseits schöpft der resultierende Abschlußtest $\bar{\varphi} = (\bar{\varphi}_1, \bar{\varphi}_2, \bar{\varphi}_{12}) := (\varphi_1\varphi_{12}, \varphi_2\varphi_{12}, \varphi_{12})$ das multiple Niveau α voll aus. □

Ähnliche Beispiele lassen sich für multiple Varianten von F–Tests in der Varianzanalyse konstruieren. Beispiel 1.1 zeigt nachdrücklich, daß das Abschlußprinzip lediglich ein Hilfsmittel zur Konstruktion von multiplen Tests ist und nicht notwendig zu einer optimalen Lösung führt. Andererseits stellt diese Methode eine einfache Möglichkeit zur Festlegung der Annahme– und Ablehnungsbereiche der tatsächlich interessierenden Hypothesen dar. Dies könnte natürlich auch ohne den Übergang zur ∩–abg. Hypothesenfamilie und zum Abschlußtest geschehen. Das Problem dabei ist nur die im allgemeinen äußerst komplizierte Struktur der Annahme– und Ablehnungsbereiche. Hier liefert das Abschlußprinzip über den Umweg der lokalen Niveau–α–Tests eine formal sehr einfache Beschreibungsmöglichkeit. Die einzige Gefahr liegt wie schon erwähnt darin, durch das Aufblähen der Hypothesenfamilie die eigentlichen Zielhypothesen aus den Augen zu verlieren.

2. Eigenschaften der Verteilungsfunktion der Spannweite

Ein wichtiges Hilfsmittel zum Nachweis, ob ein multipler Test das vorgegebene multiple Niveau einhält, sind Ungleichungen für die entsprechenden Wahrscheinlichkeitsverteilungen. Dies trifft auch für multiple Spannweitentests zu, denn aus den hier betrachteten Ungleichungen lassen sich in Kapitel 3 und Kapitel 4 nützliche Monotonieeigenschaften der entsprechenden kritischen Werte ableiten. Zunächst wird jedoch eine Formel für die Verteilungsfunktion der Spannweite hergeleitet, die in dieser allgemeinen Form in der Literatur nicht zu finden ist.

2.1. Die Verteilung der Spannweite

Seien $X_1, \ldots, X_k$ stochastisch unabhängig identisch verteilte reellwertige Zufallsvariable und $X_{(i)}$, $i = 1, \ldots, k$ die i–te Orderstatistik der X_i mit $X_{(1)} \leq \ldots \leq X_{(k)}$. Dann ist die Spannweite der X_i definiert durch

$$(2.1) \qquad R_k := R_k(X_1, \ldots, X_k) := X_{(k)} - X_{(1)} .$$

Die Verteilungsfunktion von R_k ist definiert durch

$$(2.2) \qquad \forall z \in \mathbb{R}: \; G_k(z) := P(R_k \leq z) .$$

Da diese Formel nur in seltenen Fällen nützlich ist, gilt unser Interesse zunächst alternativen Darstellungen, insbesondere solchen Formeln, die nur von der Verteilungsfunktion F der X_i abhängen. Leider sind in der Literatur nur für stetiges F explizite Formeln für G_k zu finden, während für diskrete Verteilungen in gewissen Spezialfällen die zu G_k gehörende Dichte angegeben wird (vgl. z.B. Burr (1955) oder auch David (1970), pp. 10–13). Dennoch läßt sich für beliebiges F eine recht einfache Darstellung

für G_k angeben.

Lemma 2.1. Sei D die Menge aller Unstetigkeitsstellen von F. Dann gilt:

(2.3) $\forall k \in \mathbb{N} : \forall z \geq 0 :$

$$G_k(z) = \int_{\mathbb{R}\setminus D} k\,[F(y+z) - F(y)]^{k-1}\, dF(y) + \sum_{y\in D} \{[F(y+z) - F(y^-)]^k - [F(y+z) - F(y)]^k\}$$

mit $F(y^-) := \lim_{x\uparrow y} F(x)$.

Beweis: Für $k = 1$ und $z \geq 0$ ergibt (2.3) $G_1(z) = \int_{\mathbb{R}\setminus D} dF(y) + \sum_{y\in D} [F(y) - F(y^-)] = 1.$

Aus $R_1 \equiv 0$ und $P(R_1 \leq z) = 1$ für $z \geq 0$ folgt die Behauptung.

Sei nun $k \geq 2$. Dann gilt

$$\begin{aligned}\{R_k \leq z\} &= \bigcup_{y\in\mathbb{R}} \{X_{(1)} = y\,,\, X_{(k)} \leq y + z\} \\ &= \bigcup_{y\in\mathbb{R}\setminus D} \bigcup_{i=1}^{k} \{X_i = y\,,\, \forall j \neq i :\, X_j \in [y, y+z]\} \\ &\quad + \sum_{y\in D} (\{y \leq X_{(1)}, X_{(k)} \leq y+z\} \setminus \{y < X_{(1)},\, X_{(k)} \leq y+z\}) \\ &=: A + B\,.\end{aligned}$$

Da die Wahrscheinlichkeit, daß zwei oder mehr der X_i einen gleichen Wert $y\in\mathbb{R}\setminus D$ annehmen, bekanntlich gleich 0 ist, gilt die folgende Zerlegung von A P–fast sicher:

$$\begin{aligned}A &= \bigcup_{i=1}^{k} \bigcup_{y\in\mathbb{R}\setminus D} \{X_i = y\,,\, \forall j \neq i :\, X_j \in [y, y+z]\} \\ &\overset{[P]}{=} \sum_{i=1}^{k} \bigcup_{y\in\mathbb{R}\setminus D} \{X_i = y\,,\, \forall j \neq i :\, X_j \in (y, y+z]\} \\ &=: \sum_{i=1}^{k} A_i\,.\end{aligned}$$

Damit erhält man

$$\begin{aligned} G_k(z) &= P(A) + P(B) \\ &= \sum_{i=1}^{k} P(A_i) + P(B) \\ &= k\,P(A_1) + P(B) \end{aligned}$$

mit

$$\begin{aligned} P(A_1) &= \int I_{\mathbb{R}\setminus D}(x_1) \prod_{j=2}^{k} I_{(x_1,x_1+z]}(x_j)\, dP^{(X_1,\ldots,X_k)}(x_1,\ldots,x_k) \\ &= \int_{\mathbb{R}\setminus D} [F(y+z) - F(y)]^{k-1}\, dF(y) \end{aligned}$$

und

$$P(B) = \sum_{y\in D} \{[F(y+z) - F(y^-)]^k - [F(y+z) - F(y)]^k\} \,.\ \square$$

Ist F stetig, so erhält man die bekannte Formel (vgl. z.B. Patel, Kapadia, Owen (1976), p. 64)

$$(2.4) \qquad \forall k \in \mathbb{N}: \ \forall z \geq 0: \ G_k(z) = k \int [F(y+z) - F(y)]^{k-1}\, dF(y) \,.$$

Für diskrete Verteilungen mit endlichem Träger $\{y_1,\ldots,y_m\}$ und $y_1 < \ldots < y_m$ erhält man mit $y_0 := -\infty$

$$(2.5) \qquad \forall k \in \mathbb{N}: \ \forall z \geq 0:$$

$$G_k(z) = \sum_{i=1}^{m} [F(y_i+z) - F(y_{i-1})]^k - \sum_{i=1}^{m-1} [F(y_i+z) - F(y_i)]^k \,.$$

Im allgemeinen ist die Verteilungsfunktion der Spannweite nur mit einigem numerischen Aufwand auszuwerten. Ein Beispiel dafür ist die Normalverteilung (vgl. Abschnitt 4.4). Wir betrachten hier nur einige Ausnahmen, auf die später zurückgegriffen wird.

Beispiele 2.1.

(a) Wohlbekannt ist G_k für gleichverteilte Zufallsvariable auf dem Einheitsintervall (vgl. z.B. Gumbel (1958), p. 104). Für $z \in [0,1]$ gilt:

$$(2.6) \qquad G_k(z) = k\, z^{k-1} - (k-1)\, z^k .$$

(b) Für exponentialverteilte Zufallsvariable mit der λ^1–Dichte $f_\lambda(x) = \lambda e^{-\lambda x} I_{[0,\infty)}(x)$, $\lambda > 0$, erhält man für $z \geq 0$:

$$(2.7) \qquad G_k(z) = (1 - e^{-\lambda z})^{k-1} .$$

(c) Für die diskrete Gleichverteilung auf $\{1,\ldots,m\}$ ergibt sich für $z \in \{0,1,\ldots,m\}$ die Formel

$$(2.8) \qquad G_k(z) = [(m-z)\,(z+1)^k - (m-z-1)\, z^k] \,/\, m^k .$$

2.2. Log–Konkavität und Log–Konvexität der Spannweite

Eine für abgeschlossene multiple Spannweitentests wünschenswerte Eigenschaft wird die Log–Konkavität der Verteilungsfunktion der Spannweite in k sein (vgl. Abschnitt 4.2). Hayter (1986) zeigte im Zusammenhang mit Tests im ANOVA–Modell für die Normalverteilung eine Ungleichung für G_k, die äquivalent mit der Log–Konkavität von G_k in k ist. Leider läßt sich dieses Ergebnis nicht auf beliebige Verteilungen übertragen. Dieser Abschnitt wird zeigen, daß es Verteilungen gibt, so daß G_k log–konkav, log–konvex, log–linear oder aber weder das eine noch das andere ist.

Definition 2.1. Eine Funktion $f : \mathbb{N} \to \mathbb{R}$ heißt konvex (konkav), wenn

$$(2.9) \qquad \forall\, n \in \mathbb{N}: \; 2 f(n+1) \leq (\geq)\; f(n) + f(n+2) \; .$$

Eine Funktion $f : \mathbb{N} \to (0,\infty)$ heißt log–konvex (log–konkav), wenn log f (–log f) konvex ist.
Gilt in (2.9) "<" (">"), so heißt f streng log–konvex (streng log–konkav). f heißt log–linear, falls f gleichzeitig log–konvex und log–konkav ist.

Die Log–Konvexität (Log–Konkavität) einer Funktion $f : \mathbb{N} \to (0,\infty)$ ist äquivalent mit jeder der beiden folgenden Bedingungen:

$$(2.10) \qquad \forall\, r,s \in \mathbb{N}\,,\; 2 \leq r \leq s: \; f(r)\, f(s) \leq (\geq)\; f(r-1)\, f(s+1)$$

$$(2.11) \qquad \forall\, r,s,t \in \mathbb{N}\,,\; t < r \leq s: \; f(r)\, f(s) \leq (\geq)\; f(r-t)\, f(s+t) \; .$$

Die Verteilungsfunktion der Spannweite ist also genau dann log–konvex (log–konkav) in k, falls die Bedingung

$$(2.12) \qquad \forall\, r,s \in \mathbb{N}\,,\; 2 \leq r \leq s : \forall\, z \geq 0 : \; G_r(z)\, G_s(z) \leq (\geq)\; G_{r-1}(z)\, G_{s+1}(z)$$

erfüllt ist.

Die folgende heuristische Betrachtung könnte einen zunächst glauben lassen, daß G_k stets log–konkav in k ist.
Sind $r,s \in \mathbb{N}$, $2 \leq r < s$ und X_n, $n \in \mathbb{N}$ stochastisch unabhängig identisch verteilt, so ist $G_r(z)\; G_s(z)$ mit $T_{ij} := |X_i - X_j|$, $i,j \in \mathbb{N}$ gerade die Wahrscheinlichkeit für das Ereignis

$$(2.13) \qquad A_{r,s} := \{ \max_{1 \leq i < j \leq r} T_{ij} \leq z \} \cap \{ \max_{r+1 \leq i < j \leq r+s} T_{ij} \leq z \} \; .$$

Betrachtet man den Übergang zu den Ereignissen $A_{r+1,s}$ und $A_{r,s+1}$, so stellt man fest, daß sich die Anzahl der T_{ij}, durch die diese Ereignisse bestimmt werden, für $A_{r+1,s}$ um r und für $A_{r,s+1}$ um s erhöht. Die Wahrscheinlichkeiten für die betrachteten Ereignisse werden natürlich mit jedem zusätzlichen T_{ij} kleiner. Sollte dann wegen $r < s$ nicht auch $P(A_{r+1,s}) \geq P(A_{r,s+1})$ und damit $G_{r+1}(z)\ G_s(z) \geq G_r(z)\ G_{s+1}(z)$, also die Log–Konkavität von G_k gelten? Selbst wenn man beachtet, daß die T_{ij} innerhalb der beiden Gruppen, über die die Spannweite gebildet wird, hochgradig abhängig sind, scheint diese Heuristik einer gewissen Logik nicht zu entbehren. Es gilt jedoch

Lemma 2.2. Für die b(1,p)–Verteilung mit $p \in (0,1)$, $p \neq \frac{1}{2}$ ist die Verteilungsfunktion der Spannweite log–konvex in k und für $z = 0$ streng log–konvex in k. Für $p = \frac{1}{2}$ ist G_k log–linear in k.

Beweis: Für $z \geq 1$ ist nichts zu zeigen. Ist $z \in [0,1)$, so folgt aus (2.5)
$G_k(z) = p^k + (1-p)^k$. Für $r,s \in \mathbb{N}$, $2 \leq r \leq s$, erhält man

$$
\begin{aligned}
&G_r(z)\ G_s(z) - G_{r-1}(z)\ G_{s+1}(z) \\
&\quad = (p^r + (1-p)^r)\ (p^s + (1-p)^s) - (p^{r-1} + (1-p)^{r-1})\ (p^{s+1} + (1-p)^{s+1}) \\
&\quad = p^{r-1}(1-p)^{r-1}\ (2p-1)\ [(1-p)^{s-r+1} - p^{s-r+1}] \\
&\quad \begin{cases} < 0 \ , & p \in (0,1)\ ,\ p \neq \frac{1}{2} \\ = 0 \ , & p = \frac{1}{2} \end{cases}
\end{aligned}
$$

. □

Immerhin läßt sich doch für einige gebräuchliche Verteilungen zeigen, daß G_k log–konkav ist.

Satz 2.1. (Hayter (1986)). Seien $X_k \sim N(\mu,\sigma^2)$, $k \in \mathbb{N}$, stochastisch unabhängig. Dann gilt für die Verteilungsfunktion der Spannweite der X_i

$$
(2.14) \qquad \forall\, r,s,t \in \mathbb{N}\ ,\ t < r \leq s\ :\ \forall\, z \geq 0\ :\ G_r(z)\ G_s(z) \geq G_{r-t}(z)\ G_{s+t}(z)\ ,
$$

d.h., $G_k(z)$ ist bei festem z log–konkav in k.

Der Beweis dieses Satzes beruht auf einer intelligenten Reihenentwicklung von G_k und soll an dieser Stelle nicht wiederholt werden. Es ist bisher leider nicht gelungen, diese Beweisidee auf andere Verteilungen zu übertragen.

Aufgrund der relativ einfachen Darstellungen der G_k in den Beispielen 2.1 läßt sich durch einfaches Verifizieren von (2.12) das folgende Lemma zeigen:

Lemma 2.3. Für die stetige Gleichverteilung und die diskrete Gleichverteilung mit äquidistantem Träger ist G_k log–konkav in k. Für die Exponentialverteilung mit Parameter $\lambda > 0$ ist G_k log–linear in k.

Es folgt ein Beispiel für eine Verteilung, deren Spannweite weder log–konvex noch log–konkav ist.

Beispiel 2.2. Seien X_k , $k \in \mathbb{N}$, stochastisch unabhängig identisch verteilte Zufallsvariable mit der Lebesgue–Dichte $f(x) = \frac{1}{4}(I_{[0,1]}(x) + I_{[3,4]}(x))$. Mit (2.4) erhält man für $z \in [2,3]$ die Formel

$$G_k(z) = [2 + (k-1)(z-2)^k]/2^k .$$

Für $z_0 = 5/2$ gilt zum Beispiel $G_5(z_0)\, G_5(z_0) < G_4(z_0) \cdot G_6(z_0)$, aber $G_4(z_0)\, G_4(z_0) > G_3(z_0)\, G_5(z_0)$. Das heißt, daß $G_k(z_0)$ weder log–konvex noch log–konkav in k ist. □

Abschließend sei angemerkt, daß sowohl die Log–Konvexität als auch die Log–Konkavität von G_k eine Verbesserung der elementaren Ungleichung

(2.15) $$\forall k \in \mathbb{N} : \forall z \geq 0 : G_{k+1}(z) \leq G_k(z)$$

im folgenden Sinne liefert:

Ist G_k log–konkav in k, so folgt aus (2.12) für $k \geq 2$, falls $G_{k-1}(z) \neq 0$,

$$(2.16) \qquad G_{k+1}(z) \leq \frac{G_k(z)}{G_{k-1}(z)} G_k(z) .$$

Ist G_k log–konvex in k, so folgt aus (2.12), falls $G_{k+1}(z) \neq 0$

$$(2.17) \qquad G_{k+1}(z) \leq \frac{G_{k+2}(z)}{G_{k+1}(z)} G_k(z) .$$

Wegen $\frac{G_k(z)}{G_{k-1}(z)} \leq 1$ liefern sowohl (2.16) als auch (2.17) schärfere Ungleichungen zwischen $G_{k+1}(z)$ und $G_k(z)$.

2.3. Eine grundlegende Ungleichung für die Verteilungsfunktion der Spannweite

Der folgende Satz liefert das Hauptresultat dieses Abschnitts, eine für beliebige Verteilungen gültige Verschärfung der Ungleichung (2.15), die nur von G_k, G_{k+1} und k abhängt.

Satz 2.2. Sei F eine beliebige (eindimensionale) Verteilungsfunktion. Dann gilt für die zugehörige Verteilungsfunktion der Spannweite

$$(2.18) \qquad \forall\, k \in \mathbb{N} : \forall\, z \geq 0 : G_{k+1}^{1/(k+1)}(z) \leq G_k^{1/k}(z) .$$

Zum Beweis dieses Satzes werden zunächst zwei Ungleichungen für konvexe Funktionen hergeleitet.

Lemma 2.4. Sei $h : I \to \mathbb{R}$ eine auf einem Intervall $I \subseteq \mathbb{R}$ definierte konvexe Funktion. Für Punkte $x_1,\dots,x_m, z_1,\dots,z_{m-1} \in I$ mit $m \in \mathbb{N}$, $m \geq 2$, und

$$z_i \leq \min\{x_i, x_{i+1}\},\ i=1,\dots,m-1 \quad \text{und} \quad x_m + \sum_{i=1}^{m-1}(x_i - z_i) \in I$$

gilt dann

$$\sum_{i=1}^{m} h(x_i) - \sum_{i=1}^{m-1} h(z_i) \leq h\Big(\sum_{i=1}^{m} x_i - \sum_{i=1}^{m-1} z_i\Big). \tag{2.19}$$

Beweis: Für $m = 2$ folgt (2.19) aus der Konvexität von h. Per Induktion nach m erhält man

$$\sum_{i=1}^{m+1} h(x_i) - \sum_{i=1}^{m} h(z_i) \leq h\Big(\sum_{i=1}^{m} x_i - \sum_{i=1}^{m-1} z_i\Big) + h(x_{m+1}) - h(z_m).$$

Wegen

$$z_m \leq \min\{x_m, x_{m+1}\} \leq \sum_{i=1}^{m} x_i - \sum_{i=1}^{m-1} z_i$$

folgt mit (2.19) für $m = 2$

$$h\Big(\sum_{i=1}^{m} x_i - \sum_{i=1}^{m-1} z_i\Big) + h(x_{m+1}) - h(z_m) \leq h\Big(\sum_{i=1}^{m+1} x_i - \sum_{i=1}^{m} z_i\Big). \quad \square$$

Lemma 2.5. Seien $h_j : I \to I_j \subseteq \mathbb{R}$, $j = 1,2$, auf einem Intervall $I \subseteq \mathbb{R}$ definierte bijektive, monton wachsende Funktionen, so daß $h_1 \circ h_2^{-1} : I_2 \to I_1$ konvex ist. Für Punkte $x_1,\dots,x_m, z_1,\dots,z_{m-1} \in I$ mit $m \in \mathbb{N}$, $m \geq 2$, und

$z_i \leq \min\{x_i, x_{i+1}\}$, $i=1,\dots,m-1$ und $h_j(x_m) + \sum_{i=1}^{m}(h_j(x_i) - h_j(z_i)) \in I_j$, $j = 1,2$

gilt dann

$$h_1^{-1}\Big(\sum_{i=1}^{m} h_1(x_i) - \sum_{i=1}^{m-1} h_1(z_i)\Big) \leq h_2^{-1}\Big(\sum_{i=1}^{m} h_2(x_i) - \sum_{i=1}^{m} h_2(z_i)\Big). \tag{2.20}$$

Beweis: Sei $x_i' := h_2(x_i)$, $i=1,\dots,m$ und $z_i' := h_2(z_i)$, $i=1,\dots,m-1$. Da h_2 monoton wachsend ist und $z_i \le \min\{x_i, x_{i+1}\}$, gilt $z_i' \le \min\{x_i', x_{i+1}'\}$. Zusammen mit der Konvexität von $h_1 \circ h_2^{-1}$ folgt mit Lemma 2.4 die Ungleichung

$$\sum_{i=1}^{m} h_1(h_2^{-1}(x_i')) - \sum_{i=1}^{m-1} h_1(h_2^{-1}(z_i')) \le h_1(h_2^{-1}(\sum_{i=1}^{m} x_i' - \sum_{i=1}^{m-1} z_i')) .$$

Da h_i^{-1} monoton wachsend ist, folgt (2.20). □

Beweis von Satz 2.2:

Die Ungleichung wird zunächst für diskrete Verteilungen mit endlichem Träger $\{y_1,\dots,y_m\}$, $y_1 < \dots < y_m$, bewiesen. Für $m = 1$ gilt stets $G_{k+1}^{1/(k+1)}(z) = G_k^{1/k}(z)$. Für $m \ge 2$ ist G_k mit $y_0 := -\infty$ nach (2.5) gegeben durch

$$\forall k \in \mathbb{N} : \forall z \ge 0 : \; G_k(z) = \sum_{i=1}^{m} (F(y_i+z) - F(y_{i-1}))^k - \sum_{i=1}^{m-1} F(y_i+z) - F(y_i))^k .$$

Mit
$$h_1(x) := x^{k+1} \;,\; h_2(x) := x^k \qquad , x \in [0,1] ,$$
$$x_i := F(y_i+z) - F(y_{i-1}) \qquad , i=1,\dots,m$$
$$z_i := F(y_i+z) - F(y_i) \qquad , i=1,\dots,m-1$$

folgt die behauptete Ungleichung aus Lemma 2.5.

Sei nun P^{X_i} beliebig. Dann existiert eine Folge von Wahrscheinlichkeitsmaßen $(P_\ell)_{\ell\in\mathbb{N}}$ mit endlichen Trägern (vgl. Bauer (1978), Korollar 45.4), so daß P_ℓ schwach gegen P^{X_i} konvergiert ($P_\ell \xrightarrow{w} P^{X_i}$) .

Für die Produktmaße folgt (vgl. Billingsley (1968), p. 21)

$$P_\ell \otimes \dots \otimes P_\ell \xrightarrow{w} P^{X_1} \otimes \dots \otimes P^{X_k} = P^{(X_1,\dots,X_k)} .$$

Da die Abbildung $R_k(x_1,\dots,x_k) := \max_{1\le i\le j\le k} |x_i - x_j|$ stetig und damit meßbar ist, erhält man (vgl. Billingsley (1979), Theorem 29.2)

$$(P_\ell \otimes \dots \otimes P_\ell)^{R_k} \xrightarrow{w} P^{R_k(X_1,\dots,X_k)} .$$

Folglich gilt für alle Stetigkeitspunkte z von G_k (vgl. Billingsley (1979), Theorem 25.8)

$$G_{k,\ell}(z) := (P_\ell \otimes \dots \otimes P_\ell)^{R_k}((-\infty,z]) \longrightarrow G_k(z) ,$$

also auch $G_{k+1}^{1/(k+1)}(z) \le G_k^{1/k}(z)$.

Da die Menge aller Unstetigkeitsstellen von G_k abzählbar ist, gibt es für jede Unstetigkeitsstelle z von G_k eine Folge von Stetigkeitspunkten $(z_m)_{m\in\mathbb{N}}$ mit $z_m \downarrow z$, so daß

$$G_k(z_m) \longrightarrow G_k(z) .$$

Wegen $G_{k+1}^{1/(k+1)}(z_m) \le G_k^{1/k}(z_m)$ für alle $k,m \in \mathbb{N}$, folgt mit $m \to \infty$ die Behauptung. □

Bemerkung 2.1. Aus Satz 2.2 folgt unmittelbar

(2.21) $$\forall k \in \mathbb{N} : \forall \gamma \le (k+1)/k : \forall z \in \mathbb{R} : G_{k+1}(z) \le G_k^{\gamma}(z) .$$

Andererseits existiert für jedes $\gamma > (k+1)/k$ eine Verteilungsfunktion F und ein $z \ge 0$, so daß gilt: $G_{k+1}(z) > G_k^{\gamma}(z)$. Man wähle etwa $z = 0$ und eine Bernoulli–Verteilung mit hinreichend kleinem p. In diesem Sinn ist $(k+1)/k$ eine scharfe Schranke für γ. □

Bemerkung 2.2. Mit (2.18) äquivalent ist die Formulierung

(2.22) $$\forall n,m \in \mathbb{N} , m > n : \forall z \ge 0 : G_m^n(z) \le G_n^m(z) ,$$

und die ist insofern interessant, da sie eine Schar von "trivialen" Ungleichungen enthält. Für alle $\ell, m, n \in \mathbb{N}$ mit $\ell \geq 2$ und $m = \ell n$ ist $G_m^n(z) \leq G_n^m(z)$ äquivalent mit

$$G_{\ell n}(z) \leq G_n^{\ell}(z) .$$

Mit

$$A_r^n(z) := \{ \max_{(n-1)r+1 \leq i \leq j \leq nr} |X_i - X_j| \leq z \} , \quad r = 1, \ldots, \ell$$

und

$$B(z) := \{ \max_{1 \leq i < j \leq n\ell} |X_i - X_j| \leq z \}$$

folgt diese Ungleichung schon aus der Mengeninklusion

$$B(z) \subseteq \bigcap_{r=1}^{\ell} A_r^n(z)$$

und der Unabhängigkeit der $A_r^n(z)$:

$$G_{\ell n}(z) = P(B(z)) \leq P\Big(\bigcap_{r=1}^{\ell} A_r^n(z)\Big) = \prod_{r=1}^{\ell} P(A_r^n(z)) = G_n^{\ell}(z) .$$

Der Versuch, (2.22) und damit auch Satz 2.2 vollständig auf derartige Mengeninklusionen zurückzuführen, schlägt allerdings fehl. Dann müßte nämlich $\forall\, n, m \in \mathbb{N}$, $m > n$, die Mengeninklusion

$$\bigcap_{r=1}^{m} A_r^n(z) \subseteq \bigcap_{r=1}^{n} A_r^m(z)$$

gelten. Diese gilt jedoch im allgemeinen nicht. Um ein Gegenbeispiel anzugeben, wähle man etwa $n=3$, $m=2$, $z=0$ und $X_i \sim b(1,p)$. □

Ist die Verteilungsfunktion der Spannweite log–konvex bzw. log–konkav, so läßt sich die Spannweitenungleichung (2.18) direkt aus (2.12) herleiten. Dies folgt aus einem allgemein gültigen Zusammenhang zwischen dem Ungleichungstyp

(2.23) $\forall\, k \in \mathbb{N}:\ f^{1/k+1}(k+1) \le f^{1/k}(k)$

und der Log–Konkavität bzw. Log–Konvexität von $f : \mathbb{N} \to (0,\infty)$.

Lemma 2.6. Ist $f : \mathbb{N} \to (0,\infty)$ log–konvex auf $\mathbb{N}$ mit

(2.24) $\forall\, r,s \in \mathbb{N}:\ f(2r) \le f^2(r)$,

so folgt (2.23).

Beweis: Für $i \in \mathbb{N}$ folgt aus der Log–Konvexität von f

$$f(k+1)\, f(k+i) \le f(k)\, f(k+i+1) ,$$

und damit

$$f^k(k+1) \le f^{k-2}(k+1)\, f(k)\, f(k+2) \le \ldots \le f^{k-1}(k)\, f(2k) .$$

Mit (2.24) folgt $f^k(k+1) \le f^{k+1}(k)$. □

Lemma 2.7. Ist $f : \mathbb{N} \to (0,\infty)$ log–konkav auf $\mathbb{N}$ mit

(2.25) $\forall\, r \in \mathbb{N}:\ f(r+1) \le f(r)\, f(1)$,

so folgt (2.23).

Beweis: Für $i = 0,1,\ldots,k-2$ folgt aus der Log–Konkavität von f

$$f(k)\, f(k-i) \ge f(k+1)\, f(k-i-1) ,$$

und damit

$$f^{k+1}(k) \ge f^{k-1}(k)\, f(k-1)\, f(k+1) \ge \cdots \ge f(k)\, f(1)\, f^{k-1}(k+1) .$$

Mit (2.25) folgt $f^{k+1}(k) \ge f^k(k+1)$. □

Da für G_k die Bedingungen (2.24) und (2.25) erfüllt sind, folgt die Spannweitenungleichung ummittelbar aus der Log–Konvexität bzw. Log–Konkavität von G_k.

Den Abschluß dieses Kapitels bildet eine Grenzwertaussage für die Folge $(G_k^{1/k}(z))_{k\in\mathbb{N}}$, die gleichzeitig eine untere Schranke für alle $G_k^{1/k}(z)$ liefert.

<u>Satz 2.3.</u> $$\forall\, z \geq 0 : \lim_{k\to\infty} G_k^{1/k}(z) = \sup_{x\in\mathbb{R}} [F(x+z) - F(x^-)] .$$

<u>Beweis:</u> Zunächst gilt

$$\forall\, k \in \mathbb{N}: \forall\, z \geq 0: \; G_k(z) \geq \sup_{x\in\mathbb{R}} P(X_i \in [x,x+z],\ i=1,\ldots,k) = \sup_{x\in\mathbb{R}} [F(x+z) - F(x^-)]^k .$$

Hieraus folgt

$$\lim_{k\to\infty} G_k^{1/k}(z) \geq \sup_{x\in\mathbb{R}} [F(x+z) - F(x^-)] .$$

Andererseits gilt, wenn D die Menge aller Unstetigkeitsstellen von F ist,

(i) $$\int_{\mathbb{R}\setminus D} k\,[F(x+z) - F(x)]^{k-1}\, dF(x)$$

$$\leq \int_{\mathbb{R}\setminus D} k \sup_{y\in\mathbb{R}} [F(y+z) - F(y^-)]^{k-1}\, dF(x)$$

$$= k\, P^{X_1}(\mathbb{R}\setminus D) \sup_{y\in\mathbb{R}} [F(y+z) - F(y^-)]^{k-1} ,$$

und

(ii) $$\sum_{x\in D} \{[F(x+z) - F(x^-)]^k - [F(x+z) - F(x)]^k\}$$

$$= \sum_{x\in D} P^{(X_1,\ldots,X_k)}\Big(\bigcup_{i=1}^{k} \{x_i = x\,,\ \forall\, j \neq i :\ x_j \in [x,x+z]\}\Big)$$

$$\leq \sum_{x\in D} \sum_{i=1}^{k} P^{X_i}(\{x\}) \sup_{y\in D} [F(y+z) - F(y^-)]^{k-1}$$

$$\leq k \sum_{x \in D} P^{X_1}(\{x\}) \sup_{y \in D} [F(y+z) - F(y^-)]^{k-1}$$

$$= k\, P^{X_1}(D) \sup_{y \in D} [F(y+z) - F(y^-)]^{k-1} .$$

Mit (2.3) folgt

$$G_k(z) \leq k \sup_{x \in \mathbb{R}} [F(x+z) - F(x^-)]^{k-1} ,$$

also

$$\lim_{k \to \infty} G_k^{1/k}(z) \leq \sup_{x \in \mathbb{R}} [F(x+z) - F(x^-)] . \quad \square$$

3. Klassische multiple Spannweitentests

Multiple Spannweitentests gehören wohl zu den ältesten multiplen Testverfahren. Als erster verwendete Newman (1939), inspiriert durch die Arbeiten von Galton (1902), K. Pearson (1902), E.S. Pearson (1926), Tippett (1925) und Student (1927), die (studentisierte) Spannweitenstatistik zum Prüfen von Mittelwerten bei normalverteilten Daten. Im Vergleich zu dem von Fisher (1935) vorgeschlagenen und unter dem Namen LSD–Test (least significance difference test) bekanntgewordenen Testverfahren ist der Ansatz von Newman als ein wesentlicher Fortschritt im Hinblick auf die Paarvergleiche anzusehen. Die Idee von Newman wurde später von Keuls (1952) leicht modifiziert und ist heute unter dem Namen Newman–Keuls Test bekannt. Obwohl der Newman–Keuls Test für mehr als 3 Mittelwerte kein Test zum multiplen Niveau α ist, liefert das Prinzip dieser Testprozedur nicht nur die Grundlage für dieses Kapitel, sondern auch für viele andere multiple Testprozeduren.

Ein sprunghafter Anstieg von multiplen Testverfahren für die Paarvergleiche ist Anfang der fünfziger Jahre zu verzeichnen. Hier sind insbesondere die zum Teil unveröffentlichten Arbeiten von Tukey (1949, 1952, 1953) sowie die Arbeiten von Duncan (1955) und Scheffé (1953) zu nennen. Zu diesem Zeitpunkt war allerdings keineswegs klar, welcher Niveaubegriff verwendet werden sollte. Während die Tests von Tukey und Scheffé ein multiples Niveau einhalten, ist der heuristische Ansatz von Duncan von der Einhaltung eines multiplen Niveaus noch weiter entfernt als der Newman–Keuls Test. Dennoch und vielleicht gerade deshalb erfreute sich der Duncan–Test lange Zeit großer Beliebtheit bei den Anwendern. Eine ausführliche Diskussion dieser Tests findet man in der Monographie von Miller (1966,1981) (vgl. auch O'Neill, Wetherill (1971)). Interessant ist, daß man in Miller (1966) noch keinen Hinweis auf die Arbeiten von Ryan (1959, 1960) findet, der den Newman–Keuls Spannweitentest durch die Einführung der adjustierten Signifikanzniveaus so modifizierte, daß ein multiples Niveau eingehalten wird. Das Problem einer optimalen Wahl der adjustierten Signifikanzniveaus wurde für

allgemeine Paarvergleiche sehr ausführlich in Lehmann, Shaffer (1979) (vgl. auch Lehmann (1986)) behandelt und ist auch Inhalt dieses Kapitels. Dabei wird der speziellen Struktur der Spannweitentests Rechnung getragen und es werden alternative Konzepte entwickelt. Die Bedeutung der hier betrachteten Testprozeduren, die man auch als "verkürzte" Abschlußtests bezeichnen könnnte, liegt in ihrer relativ einfachen Durchführbarkeit.

3.1. Modell, Zielhypothesen und ein klassischer Simultantest

Ausgangspunkt aller weiteren Betrachtungen ist das folgende Lokationsmodell:

Seien $X_1,\dots,X_k$ mit $k \geq 3$ stochastisch unabhängige reellwertige Zufallsvariable mit Verteilungsfunktionen

$$F(x-\mu_i) = P(X_i \leq x) \quad , i=1,\dots,k . \tag{3.1}$$

F sei bekannt und stetig und $\mu := (\mu_1,\dots,\mu_k) \in \theta = \mathbb{R}^k$.

Die Beschränkung auf stetige Lokationsfamilien ist eher formaler Natur und geschieht u.a. deshalb, um im folgenden Fallunterscheidungen zu vermeiden. Es sei jedoch angemerkt, daß sich viele der in diesem Kapitel vorgestellten Tests in kanonischer Weise auf Lokationsfamilien mit diskreten Verteilungen übertragen lassen.

Die folgenden Beispiele zeigen, daß die X_i durchaus schon Statistiken aus "größeren" Modellen, insbesondere aus k–Stichprobenmodellen mit gleichen Stichprobenumfängen sein können.

Beispiel 3.1. Seien $X_{ij} \sim N(\mu_i,\sigma^2)$, $j=1,\dots,n$, $i=1,\dots,k$ stochastisch unabhängig mit bekanntem σ^2 und unbekannten $\mu_i \in \mathbb{R}$, $i=1,\dots,k$. Definiere

$$X_i := \frac{1}{n} \sum_{j=1}^{n} X_{ij} \ , \ i=1,\ldots,k \ .$$

Beispiel 3.2. Seien X_{ij} Cauchy–verteilt, stochastisch unabhängig mit Lebesgue–Dichte $f_{ij} = [(1 + (x-\mu_i)^2)\pi]^{-1}$, $j=1,\ldots,n$, $i=1,\ldots,k$, $\mu_i \in \mathbb{R}$ unbekannt. Definiere

$$X_i := \text{Median } (X_{i1},\ldots,X_{in}) \ , \ i=1,\ldots,k \ .$$

Beispiel 3.3. Seien X_{ij} exponentialverteilt, stochastisch unabhängig mit Lebesgue–Dichte $f_{ij}(x) = \exp\{-(x-\mu_i)\}\, I_{[\mu_i,\infty)}(x)$, $j=1,\ldots,n$, $i=1,\ldots,k$, $\mu_i \in \mathbb{R}$ unbekannt. Definiere

$$X_i := \min_{1\le j\le n} X_{ij} \ , \ i=1,\ldots,k \ .$$

In allen 3 Beispielen genügen die X_i dem obigen Lokationsmodell. Der Fall ungleicher Stichprobenumfänge ist in dem hier betrachteten Modell nicht enthalten. Auch das mögliche Auftreten eines unbekannten Skalenparameters wie etwa im ANOVA–Modell wird zunächst zurückgestellt.

Zu prüfen sei, ob es Unterschiede zwischen den Parametern μ_i gibt und wo diese liegen. Als Zielhypothesen formulieren wir

$$(3.2) \qquad H_0^{ij} : \mu_i = \mu_j \ \text{ gegen } \ H_1^{ij} : \mu_i \neq \mu_j \ , \ 1 \le i < j \le k \ .$$

Die Aufgabe besteht nun in der Konstruktion eines Tests zum multiplen Niveau α für die Hypothesenfamilie $\mathcal{H} := \{H_0^{ij} : 1 \le i < j \le k\}$. Zum Prüfen dieser $\binom{k}{2}$ Hypothesen seien die Teststatistiken

$$(3.3) \qquad T_{ij} := T_{ij}(X) := |X_i - X_j| \ , \ 1 \le i < j \le k$$

mit $X := (X_1, \ldots, X_k)$ gegeben.

Unter H_0^{ij} ist die Verteilung von T_{ij} unabhängig von $\mu \in H_0^{ij}$ und gleich der Verteilung von R_2, wenn R_k , $k \in \mathbb{N}$, die Spannweite von k stochastisch unabhängig identisch nach F verteilten X_i ist. Es folgt, daß für jedes $\alpha \in (0,1)$ durch

$$\varphi_{ij} := \begin{cases} 1 \\ 0 \end{cases} , \; T_{ij} \begin{matrix} > \\ \leq \end{matrix} c_\alpha , \quad 1 \leq i < j \leq k \tag{3.4}$$

mit

$$P(R_2 \leq c_\alpha) = 1 - \alpha \tag{3.5}$$

Tests zum Niveau α für H_0^{ij} , $1 \leq i < j \leq k$, definiert sind. Mit $\mu_{ij} := |\mu_i - \mu_j|$, $1 \leq i < j \leq k$, gilt für diese Tests

$$\lim_{\mu_{ij} \to \infty} P_{\mu_{ij}}(T_{ij} > c_\alpha) = \lim_{\mu_{ij} \to \infty} E_{\mu_{ij}} \varphi_{ij} = 1 , \tag{3.6}$$

d.h., die Wahrscheinlichkeit, eine Hypothese H_0^{ij} abzulehnen, geht mit $\mu_{ij} \to \infty$ gegen 1 .

Die Tests (3.4) bzw. die Statistiken (3.3) benutzen wir als Basis zur Konstruktion von Tests zum multiplen Niveau α für $\mathcal{Y}$. Ein klassischer Lösungsansatz für dieses Testproblem ist ein Analogon zu dem aus der Varianzanalyse bekannten Tukey–Test, der auf der Spannweite der X_i bzw. dem Maximum der T_{ij} beruht.

Sei $\alpha \in (0,1)$ und $c_k(\alpha)$ bestimmt durch

$$P(R_k \leq c_k(\alpha)) = 1 - \alpha . \tag{3.7}$$

Dann ist $\varphi = (\varphi_{ij} : 1 \leq i < j \leq k)$ mit

(3.8) $$\varphi_{ij} := \begin{cases} 1 \\ 0 \end{cases} , \; T_{ij} \begin{matrix} > \\ \leq \end{matrix} c_k(\alpha) \; , \; 1 \leq i < j \leq k ,$$

ein Test zum multiplen Niveau α für $\mathcal{X}$.

Ist nämlich $\mu \in \Theta$ und $I(\mu) := \{(i,j) : \mu \in H_0^{ij}\}$ die Menge der Indizes der wahren Nullhypothesen, so gilt für die Wahrscheinlichkeit, auch nur eine wahre Nullhypothese abzulehnen

$$\begin{aligned} P_\mu \Big(\bigcup_{(i,j)\in I(\mu)} \{\varphi_{ij} = 1\} \Big) &= P_\mu \Big(\max_{(i,j)\in I(\mu)} T_{ij} > c_k(\alpha) \Big) \\ &\leq P(R_k > c_k(\alpha)) \\ &= \alpha , \end{aligned}$$

mit "=" genau dann, wenn $I(\mu) = \{(i,j) : 1 \leq i < j \leq k\}$. □

Den durch (3.8) definierten Test nennen wir im folgenden das Tukey–Analogon. Das Niveau der Einzeltests φ_{ij} ist bei diesem Simultantest für alle Paare (i,j) gleich

(3.9) $$\tilde{\alpha}_k := P(R_2 > c_k(\alpha)) = \sup_{\mu\in H_0^{ij}} P_\mu(\varphi_{ij} = 1) .$$

Für Lebesgue–stetige Verteilungen gilt für die Verteilungsfunktion der Spannweite

(3.10) $$\forall\, k \in \mathbb{N} : \forall\, z > 0 : \; [0 < G_2(z) < 1 \Rightarrow G_k(z) > G_{k+1}(z) \quad \text{und} \quad \lim_{k\to\infty} G_k(z) = 0] .$$

Hieraus folgt

(3.11) $$\alpha > \tilde{\alpha}_3 > \tilde{\alpha}_4 > \dots \quad \text{und} \quad \lim_{k\to\infty} \tilde{\alpha}_k = 0 .$$

Die letzte Ungleichungskette zeigt, daß das multiple Niveau α im Sinne der Definition 1.6.(b) nicht voll ausgeschöpft wird. Die "lokalen" Niveaus $\tilde{\alpha}_k$ streben sogar mit wachsendem k gegen 0, d.h., die Chance, Lokationsunterschiede zu erkennen, wird immer kleiner.

Ziel ist natürlich ein Test für $\mathcal{Y}$, der das multiple Niveau α voll oder doch zumindest besser ausschöpft, als der hier vorgestellte Simultantest.

Bemerkung 3.1. Eine äquivalente Formulierung der Tests (3.8) ergibt sich auf Basis der Überschreitungswahrscheinlichkeiten der Statistiken T_{ij}. Seien t_{ij} die Realisationen der T_{ij} und

$$(3.12) \qquad \forall\, 1 \le i < j \le k: \; p(t_{ij}) := P(R_2 \ge t_{ij}) .$$

Mit $\tilde{\alpha}_k$ wie oben sind die Tests (3.8) äquivalent mit

$$(3.13) \qquad \varphi_{ij} = \begin{cases} 1 \\ 0 \end{cases}, \; p(t_{ij}) \begin{matrix} < \\ \ge \end{matrix} \tilde{\alpha}_k .$$

Analog lassen sich alle anderen in dieser Arbeit betrachteten Tests umschreiben, die auf dem Maximum der T_{ij} beruhen. □

3.2 Multiple Spannweitentests auf Basis der Homogenitätshypothesen

Alle bis heute in der Literatur behandelten Verbesserungen des im letzten Abschnitts vorgestellten Simultantests beruhen auf multiplen Spannweitentests für die sogenannten Homogenitätshypothesen, die eine besonders ausgezeichnete Teilfamilie der durch $\mathcal{Y}$ erzeugten $\cap$–abg. Hypothesenfamilie $\overline{\mathcal{Y}}$ bilden.

Zur Beschreibung von $\overline{\mathcal{Y}}$ sei $I := \{1, \ldots, k\}$ und

$$\Pi_k := \bigcup_{l=1}^{k-1} \{(P_1,\ldots,P_l) : \emptyset \neq P_i \subseteq I,\ 1 \le i \le l;\ \sum_{i=1}^{l} P_i = I;$$
$$\forall\, 1 \le i < j \le l : \min_{w \in P_i} w < \min_{w \in P_j} w\}$$

die Menge aller lexikographisch geordneten Partitionen von I ohne die Partition $(\{1\},\ldots,\{k\})$. Die Anzahl der Elemente von $P_i \subseteq I$ wird mit den entsprechenden Kleinbuchstaben bezeichnet, also $p_i := |P_i|$. Analog zu Π_k definieren wir die Menge aller Partitionen von k mit Berücksichtigung der Reihenfolge und ohne die Partition $(1,\ldots,1)$ durch

$$N_k := \bigcup_{l=1}^{k-1} \{(p_1,\ldots,p_l) \in \mathbb{N}^l : \sum_{i=1}^{l} p_i = k\}.$$

Ist $P := \{i_1,\ldots,i_p\} \subseteq I$ mit $|P| \ge 2$, so heißt

$$H_0^P := \bigcap_{\substack{i,j \in P \\ i<j}} H_0^{ij} : \mu_{i_1} = \ldots = \mu_{i_p}$$

Homogenitätshypothese oder reine Schnitthypothese.

Setzt man zusätzlich $H_0^{\{i\}} : \mu \in \Theta$, so heißt jede Hypothese

$$H_0^{\pi} := H_0^{(P_1,\ldots,P_l)} := \bigcap_{i=1}^{l} H_0^{P_i}$$

mit $\pi := (P_1,\ldots,P_l) \in \Pi_k$ Partitionshypothese.

Die durch $\mathcal{H}$ erzeugte $\cap$–abg. Hypothesenfamilie $\overline{\mathcal{H}}$ ist gerade die Menge aller Partitionshypothesen, also

(3.12) $$\overline{\mathcal{H}} = \{H_0^{\pi} : \pi \in \Pi_k\}.$$

Dabei ist zu beachten, daß die durch H_0^{ij} bzw. H_0^P mit $|P| \geq 2$ beschriebenen Hypothesen spezielle Elemente von $\overline{\mathcal{H}}$ sind, die sich nur durch die Schreibweise unterscheiden. Die Zielhypothesen H_0^{ij} sind die Elementarhypothesen in $\overline{\mathcal{H}}$ und werden auch Paarhypothesen genannt. Eine spezielle Homogenitätshypothese ist die Globalhypothese

$$H_0^I : \mu_1 = \ldots = \mu_k .$$

Die Menge der Homogenitätshypothesen wird mit $\mathcal{H}$ bezeichnet, also

$$\mathcal{H} := \{H_0^P : P \subseteq I , |P| \geq 2\} . \tag{3.13}$$

Eine interessante Größe ist die Anzahl der Elemente der jeweiligen Hypothesenfamilien. Für $\mathcal{X}$ und $\mathcal{H}$ erhält man

$$|\mathcal{X}| = \binom{k}{2} \tag{3.14}$$

$$|\mathcal{H}| = 2^k - k - 1 . \tag{3.15}$$

Die Anzahl der Elemente von $\overline{\mathcal{H}}$ läßt sich mit Hilfe der Stirling Zahlen 2. Art (vgl. Abramowitz, Stegun (1968), pp. 824–5) berechnen. Für $k \geq 2$ erhält man die Formel (vgl. Shaffer (1984), p. 360):

$$|\overline{\mathcal{H}}| = \sum_{m=1}^{k-1} \frac{1}{m!} \sum_{r=1}^{m} (-1)^{m-r} \binom{m}{r} r^k \tag{3.16}$$

$$= \sum_{r=1}^{k-1} \frac{r^k}{r!} \sum_{m=0}^{k-1-r} \frac{(-1)^m}{m!} .$$

Die folgende Tabelle zeigt für einige k die Werte von $|\mathcal{X}|$, $|\mathcal{H}|$ und $|\overline{\mathcal{H}}|$:

Tabelle 3.1.

k	$\lvert\mathcal{T}\rvert$	$\lvert\hat{\mathcal{T}}\rvert$	$\lvert\bar{\mathcal{T}}\rvert$
3	3	4	4
4	6	11	14
5	10	26	51
6	15	57	202
10	45	1 013	115 974
13	78	8 178	27 644 436
20	190	1 048 555	51 724 158 235 371

Bei diesen Zahlen verwundert es nicht, daß man davor zurückschreckte, für jede der Hypothesen in $\bar{\mathcal{T}}$ einen Niveau-α-Test und anschließend auch noch den Abschlußtest durchzuführen. Es sind jedoch auch historische Gründe dafür verantwortlich, daß die meisten sequentiell verwerfenden Tests für dieses Testproblem ausschließlich auf den Homogenitätshypothesen aufgebaut wurden. Der erste multiple Test auf Basis dieser Hypothesen geht auf Newman (1939) und Keuls (1952) zurück. Allgemein lassen sich multiple Spannweitentests von diesem Typ wie folgt definieren. Zunächst wird für jede Hypothese H_0^P durch

$$\varphi_P := \begin{cases} 1 \\ 0 \end{cases} , \; T_P \begin{matrix} > \\ \leq \end{matrix} c_p(\alpha_p) \tag{3.17}$$

mit $p := |P|$, $\alpha_p \in (0,1)$ und

$$P(R_p \leq c_p(\alpha_p)) = 1 - \alpha_p \tag{3.18}$$

ein Test zum Niveau α_p definiert. Mit Hilfe dieser Tests wird nun eine Art "Abschluß" auf der Familie der Homogenitätshypothesen $\mathcal{T}$ durchgeführt. Dies ergibt den für $\mathcal{T}$ kohärenten Test $\tilde{\varphi} = (\tilde{\varphi} : P \subseteq I , |P| \geq 2)$ mit

(3.19) $$\tilde{\varphi}_p := \begin{cases} 1 , & \forall\, q \geq p : \varphi_q = 1 \\ 0 , & \text{sonst} \end{cases} .$$

Für $\alpha_p = \alpha$, $p = 2,\ldots,k$, erhält man das Analogon zum Newman–Keuls Test der Varianzanalyse. Leider ist der Newman–Keuls Test für $k \geq 4$ kein Test zum multiplen Niveau α (vgl. Hartley (1955), Begun,Gabriel (1981)). Der Grund liegt darin, daß er keine Tests zum Niveau α für die Partitionshypothesen der $\cap$–abg. Hypothesenfamilie liefert. Wählt man $\alpha_p = 1 - (1-\alpha)^{p-1}$, $p = 2,\ldots,k$, so entspricht das dem Ansatz von Duncan (1955), auf den wir hier jedoch nicht weiter eingehen werden.

Es gibt in der Literatur eine ganze Reihe von Vorschlägen, wie die α_p zu wählen sind, damit $\tilde{\varphi}$ ein Test zum multiplen Niveau α ist. In diesem Zusammenhang sind die Arbeiten von Tukey (1953), Ryan (1959/60), Einot, Gabriel (1975), Welsch (1977), Lehmann, Shaffer (1977/79) und Begun, Gabriel (1981) zu nennen. Die Idee, die α_p so zu "adjustieren", daß man insgesamt einen Test zum multiplen Niveau α erhält, wird vielfach Ryan zugeschrieben. Nach Lehmann, Shaffer (1979), p. 39, hat aber schon Tukey (1953) in einem leider bis heute unveröffentlichten Manuskript diesen Vorschlag gemacht. In jedem Fall liefert der folgende Satz die theoretische Grundlage für die Wahl der α_p.

Satz 3.1. (Vgl. Lehmann, Shaffer (1979), Theorem 1). Der durch (3.19) definierte Test $\tilde{\varphi}$ ist genau dann ein Test zum multiplen Niveau α für $\mathfrak{H}$, wenn gilt

(3.20) $$\forall\, (p_1,\ldots,p_l) \in \mathbb{N}_k : 1 - \prod_{j \in L(p_1,\ldots,p_l)} (1-\alpha_{p_j}) \leq \alpha$$

mit $L(p_1,\ldots,p_l) = \{j \in \{1,\ldots,l\} : p_j \geq 2\}$.

Beweis: Aus Satz 1.4 folgt, daß der durch

$$\forall\,(P_1,\ldots,P_l)\in\mathbb{H}_k:\ \bar{\varphi}_{(P_1,\ldots,P_l)}:=\begin{cases}1\,,\ \exists\ j\in\{1,\ldots,l\}:\tilde{\varphi}_{P_j}=1\\ 0\,,\ \text{sonst}\end{cases}$$

definierte kohärente Test $\bar{\varphi}:=(\bar{\varphi}_\pi:\pi\in\mathbb{H}_k)$ genau dann ein Test zum multiplen Niveau α für $\mathfrak{F}$ ist, wenn $\tilde{\varphi}$ ein Test zum multiplen Niveau α für $\mathfrak{F}$ ist.
Ist (3.20) erfüllt, so folgt für alle $(P_1,\ldots,P_l)\in\mathbb{H}_k$ und für alle $\mu\in\mathbb{H}_0^{(P_1,\ldots,P_l)}$:

$$\begin{aligned}P_\mu(\bar{\varphi}_{(P_1,\ldots,P_l)}=1) &= P_\mu\Big(\bigcup_{j\in L(P_1,\ldots,P_l)}\{\tilde{\varphi}_{P_j}=1\}\Big)\\ &\le P_\mu\Big(\bigcup_{j\in L(P_1,\ldots,P_l)}\{\varphi_{P_j}=1\}\Big)\\ &= 1-P_\mu\Big(\bigcap_{j\in L(P_1,\ldots,P_l)}\{\varphi_{P_j}=0\}\Big)\\ &= 1-\prod_{j\in L(P_1,\ldots,P_l)}P_\mu(\varphi_{P_j}=0)\\ &= 1-\prod_{j\in L(P_1,\ldots,P_l)}(1-\alpha_{P_j})\\ &\le \alpha\,,\end{aligned}$$

d.h., $\bar{\varphi}$ ist Test zum lokalen Niveau α. Da $\bar{\varphi}$ kohärent ist, folgt mit Satz 1.2.(b) die Behauptung.

Ist umgekehrt $\tilde{\varphi}$ Test zum multiplen Niveau α für $\mathfrak{F}$, so folgt unter Beachtung von (vgl. (3.6))

$$\forall\,H_0^P\in\mathfrak{F}:\ \sup_{\mu\in H_0^P}P_\mu(\tilde{\varphi}_P=1)=\sup_{\mu\in H_0^P}P_\mu(\varphi_P=1)=1-\alpha_P$$

die Bedingung (3.20) aus

$$\begin{aligned}\alpha &\ge \sup_{\mu\in\mathbb{H}_0^{(P_1,\ldots,P_l)}}P_\mu\Big(\bigcup_{j\in L(P_1,\ldots,P_l)}\{\tilde{\varphi}_{P_j}=1\}\Big)\\ &= \sup_{\mu\in\mathbb{H}_0^{(P_1,\ldots,P_l)}}P_\mu\Big(\bigcup_{j\in L(P_1,\ldots,P_l)}\bigcap_{Q\supseteq P_j}\{\varphi_Q=1\}\Big)\end{aligned}$$

$$= \sup_{\mu \in H_0(p_1,\ldots,p_l)} P_\mu\Big(\bigcup_{j \in L(p_1,\ldots,p_l)} \{\varphi_{p_j} = 1\}\Big)$$

$$= 1 - \prod_{j \in L(p_1,\ldots,p_l)} (1 - \alpha_{p_j}) \; . \; \square$$

Da (3.19) einige Freiheit in der Wahl der α_p läßt, stellt sich die Frage nach der optimalen Wahl der α_p. Eng verwandt mit dieser Fragestellung ist die Forderung nach der Monotonie der kritischen Werte, mit der wir uns im nächsten Abschnitt beschäftigen werden.

3.3. Das Problem der Monotonie der kritischen Werte

Wir gehen im folgenden davon aus, daß wir hauptsächlich an einem Test zum multiplen Niveau α für die Paarhypothesen interessiert sind. Die Tests für die Homogenitäts– bzw. Partitionshypothesen werden lediglich als Hilfsgrößen zur Festlegung der Annahme– und Ablehnungsbereiche für die Paarhypothesen betrachtet.

Die Forderung nach der Monotonie der kritischen Werte, also

(3.21) $$c_2(\alpha_2) \leq \ldots \leq c_k(\alpha_k) \, ,$$

ergibt sich aus der folgenden Monotonieeigenschaft der Spannweitenstatistiken:

(3.22) $$\forall \, \emptyset \neq P \subset Q \subseteq I \; : \; T_P \leq T_Q \, .$$

Wählt man nämlich

(3.23) $$c'_p := \max_{2 \leq q \leq p} c_q(\alpha_q) \quad , \quad p = 3, \ldots, k \, ,$$

so gilt einerseits $c'_2 \leq \ldots \leq c'_k$, andererseits ist der hieraus resultierende Test $(\varphi'_{ij} : 1 \leq i < j \leq k)$ für die Paarhypothesen identisch mit $(\tilde{\varphi}_{ij} : 1 \leq i < j \leq k)$, der sich aus (3.19) ergibt. Ist nämlich $\varphi'_{ij} = 0$, so existiert ein $P \subseteq I$ mit $\{i,j\} \subseteq P$ und $T_P \leq c'_p = \max_{2 \leq q \leq p} c_q(\alpha_q)$. Sei $q \in \{2,\ldots,p\}$ so, daß $c'_p = c_q(\alpha_q)$. Dann existiert ein $Q' \subseteq P$ mit $\{i,j\} \subseteq Q'$, $q' = |Q'|$ und $T_{Q'} \leq c_{q'}(\alpha_{q'})$, das heißt, $\varphi_{ij} = 0$ bzw. $\varphi_{ij} \leq \varphi'_{ij}$. Aus $c'_p \leq c_p(\alpha_p)$, $p = 2,\ldots,k$, folgt umgekehrt $\varphi_{ij} \geq \varphi'_{ij}$, also insgesamt $\varphi_{ij} = \varphi'_{ij}$ für alle $1 \leq i < j \leq k$.

Sind die kritischen Werte monoton fallend, so ist der entsprechende Test für $\mathfrak{X}$ bzw. der im Beweis von Satz 3.1 konstruierte Test für $\overline{\mathfrak{X}}$ lokal konsonant. Diese Eigenschaft wird allerdings beim Übergang zu den kritischen Werten (3.23) durch Verkleinerung der Niveaus für die Homogenitätshypothesen erkauft. Diese sind jetzt gegeben durch

$$\alpha'_p := P(R_p > c'_p) \ , \ p=2,\ldots,k \ . \tag{3.24}$$

Da sich die Testentscheidungen für die Paarhypothesen beim Übergang von den kritischen Werten c_p nach c'_p bzw. von den adjustierten Signifikanzniveaus α_p nach α'_p nicht ändern, spielt ein eventueller Verlust hinsichtlich des Niveaus für die Homogenitätshypothesen keine Rolle. Die Forderung (3.21) hat zudem den Vorteil, daß man auch ohne Berechnung aller Test $\tilde{\varphi} = (\tilde{\varphi}_P : P \subseteq I \ , \ |P| \geq 2)$ den resultierenden Test $(\tilde{\varphi}_{ij} : 1 \leq i < j \leq k)$ für die Paarhypothesen berechnen kann.

Bezeichnet $X_{(i)}$ wieder die i–te Orderstatistik der X_i mit $X_{(1)} \leq \ldots \leq X_{(k)}$, und ist $t_{ij} := X_{(j)} - X_{(i)}$, $1 \leq i < j \leq k$, so ist der Test $\psi := (\psi_{ij} : 1 \leq i < j \leq k)$ mit

$$\psi_{ij} = \begin{cases} 1 \ , & \forall \ 1 \leq r \leq i < j \leq s \leq k : t_{rs} > c_{s-r+1}(\alpha_{s-r+1}) \\ 0 \ , & \text{sonst} \end{cases} \tag{3.25}$$

nach entsprechender Indextransformation identisch mit $(\tilde{\varphi}_{ij} : 1 \le i < j \le k)$. Schematisch läßt sich die Struktur der durch (3.25) gegebenen Testprozedur wie folgt veranschaulichen:

Tabelle 3.2:

Stufe	kritischer Vert	Hypothesen
1	$c_k(\alpha_k)$	$(1,k)$
2	$c_{k-1}(\alpha_{k-1})$	$(1,k-1)$ $(2,k)$
3	$c_{k-2}(\alpha_{k-2})$	$(1,k-2)$ $(2,k-1)$ $(3,k)$
⋮	⋮	⋮
k-1	$c_2(\alpha_2)$	$(1,2)$ $(2,3)$ ... $(k-1,k)$

Ist für ein Paar (i,j) $\psi_{ij} = 0$, so folgt $\psi_{rs} = 0$ für alle (r,s) mit $i \le r < s \le j$.

Zu beachten ist, daß die Tests $(\tilde{\varphi}_{ij} : 1 \le i < j \le k)$ und $\psi = (\psi_{ij} : 1 \le i < j \le k)$ nicht mehr notwendig die gleichen Testergebnisse liefern, falls die Monotonieeigenschaft (3.21) verletzt ist. In solchen Fällen ist auch nicht mehr klar, ob ψ ein Test zum multiplen Niveau α ist. Im Programmpaket SAS wird zum Beispiel in den Prozeduren ANOVA und GLM unter der Option REGWQ ein Test vom Typ (3.25) verwendet, obwohl (3.21) häufig verletzt ist.

An dieser Stelle sei darauf hingewiesen, daß die Forderung nach der Monotonie der kritischen Werte für Tests vom Typ (3.25) schon von Lehmann, Shaffer (1977) aufgestellt wurde. Sie betrachteten allerdings kein multiples Niveau, sondern die Wahrscheinlichkeit, bei einer vorgegebenen Konfiguration von μ_i's irgendeinen Fehler 1. Art zu begehen. Definiert man für $(P_1,\ldots,P_l) \in \Pi_k$

$$(3.26) \qquad \alpha(P_1,\ldots,P_l) := \sup_{\mu \in \Pi_0(P_1,\ldots,P_l)} P_\mu \Big(\bigcup_{(r,s)\in J} \{\psi_{rs} = 1\} \Big)$$

mit $J := J(P_1,\ldots,P_l) := \{(i,j) : \exists\, m \in \{1,\ldots,l\} : \{i,j\} \subseteq P_m\}$, so gilt der folgende (vgl. Satz 3.1)

Satz 3.2. (Lehmann, Shaffer (1977)). Sei $\psi = (\psi_{ij} : 1 \le i < j \le k)$ wie in (3.25) definiert. Dann gilt:
(3.21) ist genau dann erfüllt, wenn

$$(3.27) \qquad \forall\, (P_1,\ldots,P_l) \in \Pi_k : \alpha(P_1,\ldots,P_l) = 1 - \prod_{i=1}^{l} (1-\alpha_{p_i}) .$$

Dabei wird $\alpha_1 := 0$ gesetzt.

Bemerkung 3.2. Aus Satz 3.2 folgt, daß die Bedingung (3.20) aus Satz 3.1 zusammen mit (3.21) hinreichend dafür ist, daß ψ ein Test zum multiplen Niveau α ist. □

Wir werden im folgenden nur noch Tests vom Typ (3.25) betrachten, deren kritische Werte monoton fallend sind. Deshalb definieren wir für $\alpha \in (0,1)$

Ψ_α := Menge der durch (3.25) definierten Tests, deren kritische Werte (3.21) erfüllen und die Tests zum multiplen Niveau α für $\mathcal{X}$ sind.

Nach welchen Kriterien ein Test $\psi \in \Psi_\alpha$ zu wählen ist, wird in den nächsten beiden Abschnitten untersucht.

3.4. Eine universelle Wahl der adjustierten Signifikanzniveaus

Es ist intuitiv klar, daß man eine Paarhypothese um so eher ablehnen kann, je kleiner man die kritischen Werte wählt. Das aber ist äquivalent dazu, die α_p möglichst

groß zu wählen. Da eine simultane Maximierung der a_p im allgemeinen nicht möglich ist, wurde von Lehmann, Shaffer (1979) in einem etwas allgemeineren Kontext der Begriff der zulässigen Tests eingeführt. Diese Definition wird hier insofern erweitert, daß die Bedingung (3.21), also die Monotonie der kritischen Werte, implizit mitgefordert wird.

Definition 3.1. (Vgl. Lehmann, Shaffer (1979), p. 35). Ein Test $\psi \in \Phi_\alpha$ mit den Niveaus $(a_2,\ldots,a_k)$ heißt zulässig, falls

(3.28) $\forall\, \psi' \in \Phi_\alpha$ mit Niveaus $(a_2',\ldots,a_k')$ gilt:
$[\exists\, i \in \{2,\ldots,k\} : a_i < a_i' \Rightarrow \exists\, j \in \{2,\ldots,k\} : a_j > a_j']$.

Lemma 3.1. (Vgl. Lehmann, Shaffer (1979), Theorem 2) Ist $\psi \in \Phi_\alpha$ mit den Niveaus $(a_2,\ldots,a_k)$ zulässig, dann gilt

(3.29) $a_2 \le \ldots \le a_k \le \alpha$.

Beweis: Die Behauptung folgt aus der Monotonie der kritischen Werte und der Monotonie von G_k in k. □

Damit ein Test ψ mit den Niveaus $(a_2,\ldots,a_k)$ überhaupt ein Test zum multiplen Niveau α ist, muß nach Satz 3.1 und Satz 3.2 die Bedingung (3.20), also

$$\forall\, (p_1,\ldots,p_l) \in N_k : \; 1 - \prod_{j \in L(p_1,\ldots,p_l)} (1 - a_{p_j}) \le \alpha$$

mit $L(p_1,\ldots,p_l) := \{j \in \{1,\ldots,l\} : p_j \ge 2\}$ erfüllt sein.
Eine heute noch gebräuchliche Wahl der a_p , die wohl schon auf Tukey (1953) zurückgeht (vgl. Lehmann, Shaffer (1979)) und die die Bedingung (3.20) erfüllt, ist

gegeben durch

$$(3.30) \qquad \alpha_p = \begin{cases} \alpha & , \ p=k-1,k \\ 1-(1-\alpha)^{p/k} & , \ p=2,\ldots,k-2 \ . \end{cases}$$

Es läßt sich leicht nachprüfen, daß sich diese α_p nicht gleichmäßig vergrößern lassen, ohne dadurch (3.20) aus Satz 3.1 und damit das multiple Niveau α zu verletzen. Bei dieser Wahl kann jedoch die Monotonieeigenschaft (3.21) verletzt sein:

Beispiel 3.4. Habe F die λ^1–Dichte $f(x) = \eta\, I_{(0,1)}(x) + (1-\eta)\, I_{(3,4)}(x)$ mit $\eta = 1/2 + ((1-\alpha)^{1/2}/2 - 1/4)^{1/2}$. Für $k = 4$, $\alpha = 0.05$ und α_p wie in (3.30) erhält man mit Hilfe der Formel (2.4) $c_2(\alpha_2) = 1 > c_3(\alpha_3) = 0.934$, also eine Verletzung von (3.21). □

Der folgende Satz zeigt, daß sich durch eine kleine Modifikation von (3.30) eine universelle Wahl der α_p ergibt, die unabhängig von der Ausgangsverteilung F die Monotonie der kritischen Werte garantiert. Der Beweis ist eine direkte Anwendung der Spannweitenungleichung (2.18) in Satz 2.2.

Satz 3.3. Sei F eine beliebige (eindimensionale) Verteilungsfunktion (nicht notwendig stetig!) und $\alpha \in (0,1)$. Dann gilt für die zu

$$(3.31) \qquad \alpha_p = 1-(1-\alpha)^{p/k} \quad , \ p = 2,\ldots,k$$

gehörenden kritischen Werte der Spannweite

$$(3.32) \qquad c_p(\alpha_p) := \inf\{z : G_p(z) \geq 1-\alpha_p\}$$

die Monotonieeigenschaft (3.21).

Beweis: Sei $p \in \{2,\ldots,k-1\}$. Aus der Definition von a_p, c_p und G_p sowie (2.18) aus Satz 2.2 folgt

$$(1-a)^{(p+1)/k} = 1 - a_{p+1} \le G_{p+1}(c_{p+1}(a_{p+1})) \le G_p^{(p+1)/p}(c_{p+1}(a_{p+1})) ,$$

also

$$1 - a_p = (1-a)^{p/k} \le G_p(c_{p+1}(a_{p+1})) .$$

Aus der Definition von $c_p(a_p)$ und der letzten Ungleichung folgt schließlich $c_p(a_p) \le c_{p+1}(a_{p+1})$. □

Die a_p aus (3.31) wurden zum Beispiel von Einot, Gabriel (1975) und Begun, Gabriel (1981) benutzt. Ein Vergleich der Werte von Tukey in (3.30) mit (3.31) zeigt, daß sich nur a_{k-1} verändert hat. Der durch (3.31) definierte Test ist zwar immer aus Φ_a, jedoch im allgemeinen nicht zulässig, da (3.28) verletzt sein kann. Um eine zulässige Wahl zu erhalten, ist a_{k-1} zwischen a und $1 - (1-a)^{(k-1)/k}$ möglichst groß zu wählen, ohne (3.21) zu verletzen. Ein Nachteil dieses Verfahrens ist die Abhängigkeit von a_{k-1} von der Ausgangsverteilung F.

Lemma 3.2 Sei F stetig. Dann ist der durch

$$(3.33)\quad a_p := \begin{cases} a & , p=k \\ P(R_{k-1} > \max\{c_{k-1}(a), c_{k-2}(1-(1-a)^{(k-2)/k})\}) & , p=k-1 \\ 1 - (1-a)^{p/k} & , p=2,\ldots,k-2 \end{cases}$$

definierte Test $\varphi \in \Phi_a$ zulässig.

Da die a_p aus (3.33) bis auf a_{k-1} nicht von F abhängen, nennen wir die resultierende Testprozedur im folgenden die natürliche Prozedur.

Bemerkung 3.3. Für $k = 3$ ist $a_2 = a_3 = a$ die einzige zulässige Wahl (vgl. Lehmann, Shaffer (1979)). Der resultierende Test ist identisch mit dem Newman–Keuls Analogon. Im Fall $k = 4$ ist (3.33) ebenfalls die einzige zulässige Wahl. Ist die Verteilungsfunktion der Spannweite G_k log–konkav in k, so stimmt (3.33) für $k = 4$ mit (3.30) überein. □

Für $k \geq 5$ gibt es im allgemeinen mehrere zulässige Lösungen. Deshalb werden im nächsten Abschnitt zwei Alternativen zu (3.33) vorgestellt.

3.5. Minimax– und Maximax–Prozeduren

Da es für $k \geq 5$ im allgemeinen keine eindeutig bestimmte Wahl der a_p gibt, stellt sich die Frage nach zusätzlichen Optimalitätsforderungen an die a_p. So führten Lehmann, Shaffer (1979) die minimax–optimalen Tests ein. Dabei werden die a_p so gewählt, daß das Minimum aller a_p, also a_2, maximiert wird. Dieser Ansatz ist insofern naheliegend, als man die Niveaus für die Paarhypothesen bzw. $a(P)$, $|P| = 2$, maximieren möchte. Für $k \geq 8$, k gerade, führt diese Methode jedoch nicht mehr zu einer eindeutigen Lösung. Aus diesem Grund wird das Minimax–Konzept hier so erweitert, daß die a_p sequentiell maximiert werden. Wegen Bemerkung 3.3 wird im folgenden stets $k \geq 5$ vorausgesetzt.

Definition 3.2. Ein Test $\varphi \in \Phi_a$ mit den Niveaus $(a_2, \ldots, a_k)$ heißt minimax–optimal, falls

(a) φ zulässig ist und

(b) $\forall\, \varphi' \in \Phi_a$ mit den Niveaus $(a_2', \ldots, a_k')$ gilt

(i) $a_2 \geq a_2'$

(ii) $[\exists\, i \in \{3, \ldots, k\} : a_i' > a_i \Rightarrow \exists\, j \in \{2, \ldots, i-1\} : a_j > a_j']$.

Für gerades k erhält man, daß die natürliche Wahl (3.33) minimax–optimal ist. Dies entspricht im wesentlichen der von Lehmann, Shaffer (1979) angegebenen Kompromißwahl. Für ungerades k wählt man zunächst die im Sinne von Lehmann, Shaffer (1979), Theorem 4, minimax–optimalen a_p (dabei wird die Monotonie der kritischen Werte nicht gefordert!)

$$(3.34) \qquad a_p := \begin{cases} a & , p=k-1,k \\ 1-(1-a)^{p/(k-1)} & , p=2,\ldots,k-2 \text{ , p gerade} \\ 1-(1-a)^{(p-1)/(k-1)} & , p=2,\ldots,k-2 \text{ , p ungerade .} \end{cases}$$

Die resultierenden kritischen Werte werden anschließend, falls nötig, gemäß (3.23) korrigiert, um die Monotonie zu erzwingen. Der Test mit den neuen, durch

$$c_p' := \max_{2\le q\le p} c_q(a_q) \quad , \quad p=3,\ldots,k$$

definierten kritischen Werten ist dann minimax–optimal. Ein Nachteil dieser Prozedur ist die große Abhängigkeit von der Ausgangsverteilung F.

Da eine Hypothese H_0^{ij} nur dann abgelehnt wird, wenn vorher alle nach den Werten der entsprechenden Statistiken geordneten Hypothesen H_0^{rs} mit $1 \le r \le i < j \le s \le k$, $(r,s) \neq (i,j)$, abgelehnt worden sind, ist es ebenso naheliegend, die kritischen Werte auf den ersten Stufen möglichst klein und damit die entsprechenden a_p möglichst groß zu wählen. Dies führt in Analogie zur Definition 3.2 zu einem Maximax–Ansatz.

Definition 3.3. Ein Test $\varphi \in \Phi_a$ mit den Niveaus $(a_2,\ldots,a_k)$ heißt maximax–optimal, falls

(a) φ zulässig ist und

(b) $\forall\, \varphi' \in \Phi_a$ mit den Niveaus $(a_2',\ldots,a_k')$ gilt

(i) $a_k \ge a_k'$

(ii) $[\exists\, i \in \{2,\ldots,k-1\} : a_i' > a_i \Rightarrow \exists\, j \in \{i+1,\ldots,k\} : a_j > a_j']$.

Wie beim Minimax–Ansatz ist auch hier der maximax–optimale Test eindeutig bestimmt. Zunächst werden die kritischen Werte berechnet. Es ist klar, daß man $a_k = a$ und damit $c_k(a_k) = c_k(a)$ wählen muß. Für $p \in \{2,\dots,k-2\}$ seien R_p und R_{k-p} die Spannweiten von p und k–p stochastisch unabhängig identisch nach F verteilten Zufallsvariablen. Definiere

$$(3.35)\quad c_{k-p,p}(a) := \inf\{c \in \mathbb{R} : P(\max\{R_p, R_{k-p}\} \le c) \ge 1-a\}\ ,\ 2 \le p \le k/2\ .$$

Da F als stetig vorausgesetzt wurde, ist $c_{k-p,p}(a)$ durch die Gleichung $P(\max\{R_p, R_{k-p}\} \le c) = 1 - a$ bestimmt. Setzt man

$$q := \min\{p : c_{k-p,p} = \max_{2\le\ell\le k/2} c_{k-\ell,\ell}\}\ ,$$

so werden durch

$$(3.36)\quad c_p := \begin{cases} c_k(a) & ,\ p = k \\ \max\{c_{k-1}(a)\ ,\ c_{k-q,q}(a)\} & ,\ p = k\text{-}1 \\ c_{k-q,q}(a) & ,\ q \le p \le k-2\ . \end{cases}$$

die ersten $k - q + 1$ kritischen Werte der Maximax–Prozedur festgelegt. Die entsprechenden Niveaus sind gegeben durch

$$(3.37)\quad a_p = \begin{cases} a & ,\ p = k \\ P(R_p > c_p) & ,\ q \le p \le k-1\ . \end{cases}$$

Ist $q > 2$, so werden die verbleibenden a_p , $2 \le p < q$ durch sequentielles Minimieren der entsprechenden c_p festgelegt. Dazu ermittelt man zunächst das größte c mit $c \le c_q$, so daß die zu den kritischen Werten

$$(3.38)\quad \tilde{c}_p = \begin{cases} c_p & ,\ q \le p \le k \\ c & ,\ 2 \le p \le q-1 \end{cases}$$

gehörenden Niveaus $\tilde{a}_p := P(G_p > \tilde{c}_p)$, $p=2,\ldots,k$, die Bedingung (3.20) erfüllen. Damit ist c_{q-1} bzw. a_{q-1} der Maximax–Prozedur durch $c_{q-1} := c$ bzw. $a_{q-1} := \tilde{a}_{q-1}$ festgelegt. Dieses Verfahren wird so lange fortgesetzt, bis alle c_p bzw. a_p festgelegt sind.

Bemerkung 3.4. Ist die Verteilungsfunktion der Spannweite log–konkav in k (vgl. Abschnitt 2.2), so sind die kritischen Werte der Maximax–Prozedur gegeben durch

$$(3.39) \qquad c_p = \begin{cases} c_p(a) & , \ p = k-1,k \\ c_{k-2,2}(a) & , \ p = 2,\ldots,k-2 \ . \end{cases}$$

Da es maximal 3 verschiedene kritische Werte gibt, reduziert sich der entsprechende Test zu einem dreistufigen multiplen Spannweitentest. Das folgende Beispiel zeigt, daß sich die Minimax– und Maximax–Prozedur sehr stark voneinander unterscheiden können. Ein guter Kompromiß zwischen diesen beiden extremen Verfahren scheint die durch (3.33) definierte natürliche Prozedur zu sein. □

Beispiel 3.5. Für die Exponentialverteilung mit Parameter $\lambda = 1$ ist die Verteilungsfunktion der Spannweite nach Beispiel 2.1(b) gegeben durch

$$G_k(z) = (1-e^{-z})^{k-1} I_{(0,\infty)}(z) \ .$$

Da G_k log–linear in k ist, gilt $c_{k-1}(a) = c_{k-2,2}(a)$. Damit reduziert sich der Maximax–Test zu einem zweistufigen multiplen Spannweitentest. Für $k = 7$ und $a = 0.05$ erhält man folgende kritische Werte und Niveaus für die (a) Minimax– (b) natürliche und (c) Maximax–Prozedur:

Tabelle 3.3:

	$100\ \alpha_p$			$c_p(\alpha_p)$		
p	(a)	(b)	(c)	(a)	(b)	(c)
7	5.00	5.00	5.00	4.77	4.77	4.77
6	4.18	4.48	5.00	4.77	4.70	4.59
5	3.36	3.60	4.02	4.77	4.70	4.59
4	2.53	2.89	3.03	4.77	4.63	4.59
3	1.70	2.17	2.03	4.77	4.52	4.59
2	1.70	1.45	1.02	4.08	4.23	4.59

Hier wird sehr deutlich, daß der günstige Wert von c_2 bei der Minimax–Prozedur auf Kosten aller vorherigen Stufen geht. Bei der Maximax–Prozedur hat man zwar den besten Wert auf der 2. Stufe, auf den weiteren Stufen jedoch keine Verbesserung mehr. Bei der natürlichen Prozedur hingegen wird der durch das sequentielle Vorgehen erzielte Gewinn relativ gleichmäßig auf die einzelnen Stufen verteilt. □

Bemerkung 3.5. Für $k \geq 4$ schöpft keine der in den letzten Abschnitten vorgestellten Testprozeduren das multiple Niveau α voll aus. Insbesondere ist für die Paarhypothesen das maximal erreichbare Niveau gegeben durch

$$\alpha(\{i,j\}) = \alpha_2$$

und strebt mit $k \to \infty$ gegen 0. □

Naturgemäß ist α_2 bei der Minimax–Prozedur am größten. Hieraus ist jedoch keineswegs der Schluß zu ziehen, daß die Minimax–Prozedur den anderen Prozeduren gleichmäßig überlegen ist. Sind etwa in Beispiel 3.5 die Realisationen der entsprechenden Zufallsvariablen gegeben durch $x_1 = 1.0$, $x_2 = \ldots = x_6 = 5.6$ und $x_7 = 6.0$, so erhält man für (a) und (b) lediglich $\psi_{17} = 1$, für (c) jedoch $\psi_{12} = \ldots = \psi_{17} = 1$, also deutlich mehr Ablehnungen als in (a) und (b). Ebenso lassen sich Zahlenbeispiele konstruieren, in denen (a) bzw. (b) mehr Ablehnungen liefern als die

anderen Prozeduren. Die Minimax–Prozedur ist wohl dann gut, wenn viele Paarhypothesen "falsch" sind, während die Maximax–Prozedur eher besser sein dürfte, wenn wenige Paarhypothesen "falsch" sind.

Bemerkung 3.6. Die in diesem Kapitel erzielten Resultate lassen sich nur mit Einschränkungen auf Lokationsmodelle mit einem zusätzlichen unbekannten Skalenparameter übertragen. Das wichtigste Beispiel ist das ANOVA–Modell mit unbekannter Varianz σ^2, das zur Verwendung von studentisierten Spannweitenstatistiken führt. Unglücklicherweise gilt Satz 2.2 im allgemeinen nicht für die Verteilungsfunktion der studentisierten Spannweite. Das ist eine Konsequenz daraus, daß die studentisierten Stichprobenmittel positiv abhängig sind (vgl. Lehmann (1966) oder Miller (1981), p. 101). Im Fall der Normalverteilung erhält man z.B. für $k = 10$, $\alpha = 0.1$ und α_p aus (3.31) für die kritischen Werte der studentisierten Spannweite mit $\nu = 2$ Freiheitsgraden

$$q_{10,2,\alpha} < q_{p,2,\alpha_p} \quad \text{für} \quad p = 2,\ldots,9\,,$$

d.h., die Monotonieeigenschaft (3.21) geht im Gegensatz zu Satz 3.3 verloren. Für nicht zu kleine Werte von ν scheint die Monotonie der kritischen Werte allerdings erhalten zu bleiben, so daß man in diesen Fällen die natürliche Prozedur verwenden kann. Mit Hilfe der Kimball–Ungleichung (Kimball (1951)) bzw. der verbesserten Bonferroni–Ungleichung (Miller (1981), p. 101) läßt sich bekanntlich die Einhaltung des multiplen Niveaus leicht nachweisen (vgl. z.B. Einot, Gabriel (1975)). Durch die Verwendung dieser Abschätzungen geht jedoch die Zulässigkeit der Testprozedur verloren. Etwas anders liegt der Fall bei der Maximax–Prozedur. Wie wir in Kapitel 4 noch sehen werden, lassen sich die kritischen Werte $c_{p,k-p}(\alpha)$ der entsprechenden studentisierten Statistik exakt berechnen. Da die Verteilungsfunktion der Spannweite der Normalverteilung log–konkav in k ist, benötigt man nur die Werte $c_k(\alpha)$, $c_{k-1}(\alpha)$ und $c_{2,k-2}(\alpha)$ für ν Freiheitsgrade zur Bestimmung der Maximax–Prozedur. Da in diesem Fall keine Abschätzungen verwendet werden, bleibt die Maximax–Prozedur zulässig, was ein gewisser Vorteil dieses Ansatzes ist.□

4. Abgeschlossene multiple Spannweitentests

Ziel dieses Kapitels ist ein Abschluß des Newman–Keuls Tests, der im Gegensatz zu allen anderen bisher bekannten Tests das multiple Niveau α voll ausschöpft. Dabei liegt das Problem weniger in der Definition eines solchen Tests als in der praktischen Durchführung. So wird in Abschnitt 4.2 zunächst in einer Reihe von Sätzen gezeigt, wie sich die Anzahl der zu prüfenden Hypothesen des hier betrachteten Abschlußtests verringern läßt. Die nach Bemerkung 4.2 aufgeworfene Frage und deren negative Beantwortung durch das überraschende Beispiel 4.1 bewog unter anderem dazu, ein strenges Satz–Beweis–Schema einzuhalten. Durch dieses Beispiel und die nachfolgenden Sätze wird jedoch auch der Einfluß der im allgemeinen nicht gültigen Log–Konkavität der Verteilungsfunktionen der Spannweite in k auf die möglichen Entscheidungsmuster deutlich. In Abschnitt 4.3 wird in groben Zügen ein Algorithmus zur Bestimmung des Abschlußtests dargestellt, der auf den Ergebnissen von Abschnitt 4.2 beruht. Das für die Praxis wichtige ANOVA–Modell wird in Abschnitt 4.4 behandelt. Dabei wirft der unbekannte Skalenparameter σ^2 keine wesentlich neuen Probleme auf, so daß alle Ergebnisse auf diese Situation übertragen werden können. An dem schon in der Einleitung vorgestellten Beispiel von Keuls (1952) mit 13 Mittelwerten wird die Arbeitsweise der Testprozedur demonstriert.

4.1. Wahl der lokalen Tests

Wir gehen wie in Kapitel 3 von einem Lokationsmodell aus, ohne dabei die Stetigkeit der Ausgangsverteilung F zu fordern. Seien $X_1, \ldots, X_k$ mit $k \geq 3$ stochastisch unabhängige reellwertige Zufallsvariable mit Verteilungsfunktionen

$$(4.1) \qquad F(x-\mu_i) = P(X_i \leq x) \quad , \quad i = 1, \ldots, k \; .$$

F sei bekannt (nicht notwendig stetig) und $\mu \in \theta \subseteq \mathbb{R}^k$. Ist F stetig, so wird $\theta = \mathbb{R}^k$ gewählt, anderenfalls ist θ in Einklang mit dem Träger der Ausgangsverteilung zu spezifizieren (z.B. $\theta = \mathbb{Z}^k$, falls der Träger gleich $\mathbb{Z}$ ist). Wir stellen uns die Aufgabe, für alle Paarvergleiche der Lokationsparameter

$$(4.2) \qquad H_0^{ij} : \mu_i = \mu_j \quad \text{gegen} \quad H_1^{ij} : \mu_i \neq \mu_j \,, \; 1 \leq i < j \leq k$$

mit Hilfe des in Abschnitt 1.2 vorgestellten Abschlußprinzips unter Verwendung von Spannweitentests einen Test zum multiplen Niveau α , $\alpha \in (0,1)$, zu konstruieren. Wie in Kapitel 3 (vgl. Abschnitt 3.2) bezeichnen wir die Menge der Paarhypothesen mit $\mathcal{Y}$, die Menge der Homogenitätshypothesen mit $\mathcal{H}$ und die $\cap$–abg. Hypothesenfamilie mit $\overline{\mathcal{H}}$. Die Menge aller Partitionen von k wird wieder mit N_k und die Menge aller Partitionen von $I = \{1,\ldots,k\}$ mit Π_k bezeichnet.

Bevor man das Abschlußprinzip anwenden kann, sind zunächst Niveau–α–Tests für alle Hypothesen $H_0^\tau \in \overline{\mathcal{H}}$ mit $\tau = (P_1,\ldots,P_l) \in \Pi_k$ festzulegen. Da wir uns auf Spannweitentests beschränken wollen, wählen wir für die Homogenitätshypothesen $H_0^P \in \mathcal{H}$ wie in Kapitel 3 die Tests

$$(4.3) \qquad \varphi_P(X) := \begin{cases} 1 \\ 0 \end{cases} \,, \; T_P(X) \begin{matrix} > \\ \leq \end{matrix} c_\alpha(p)$$

mit $X := (X_1,\ldots,X_k)$, $p := |P|$ und

$$(4.4) \qquad c_\alpha(p) := \inf \{c : P(R_p \leq c) \geq 1 - \alpha\} \,,$$

wobei R_p die Spannweite von p stochastisch unabhängig identisch nach F verteilten X_i ist. Da sich alle Partitionshypothesen als Durchschnitt von Homogenitätshypothesen bzw. Paarhypothesen ergeben, kann man zur Konstruktion von Tests für die Partitionshypothesen das Vereinigungs–Durchschnittsprinzip (Roy (1953), Roy,Bose (1953))

anwenden (vgl. Sonnemann (1982)). Dabei ist nur die "Gewichtung" der einzelnen Homogenitäts– bzw. Paarhypothesen mit $H_0^P \supseteq H_0^{(P_1,\ldots,P_l)}$ bzw. $H_0^{ij} \supseteq H_0^{(P_1,\ldots,P_l)}$ festzulegen.

Peritz (1970) schlug vor, die Tests für die Partitionshypothesen auf den in Kapitel 3 vorgestellten Signifikanzniveaus aufzubauen und für die Homogenitätshypothesen die Niveau–α–Tests (4.3) zu verwenden. Wählt man z.B. die adjustierten Signifikanzniveaus (3.30) von Tukey

$$(4.5) \qquad \alpha_p = \begin{cases} \alpha & , \ p = k-1,k \\ 1-(1-\alpha)^{p/k} & , \ p = 2,\ldots,k-2 \ , \end{cases}$$

so führt das für die Partitionshypothesen $H_0^{(P_1,\ldots,P_l)} \in \mathbb{H}_k$, die nicht Homogenitätshypothesen sind, zu den Tests

$$(4.6) \qquad \varphi_{(P_1,\ldots,P_l)} = \begin{cases} 1 \ , \ \exists \ i \in \{1,\ldots,l\} \ , \ p_i \geq 2 : T_{p_i} > c_{p_i}(\alpha_{p_i}) \\ 0 \ , \ \text{sonst} \end{cases} .$$

Der hieraus resultierende Abschlußtest $\bar{\varphi} := (\bar{\varphi}_\pi : \ \pi \in \mathbb{H}_k)$ mit

$$(4.7) \qquad \bar{\varphi}_\pi := \begin{cases} 1 \ , \ \forall \ H_0^{\pi'} \subseteq H_0^\pi : \ \varphi_{\pi'} = 1 \\ 0 \ , \ \text{sonst} \end{cases}$$

ist dann nach Satz 1.2 ein Test zum multiplen Niveau α. Begun, Gabriel (1981) geben für die Durchführung dieser Testprozedur einen praktikablen Algorithmus an. Simulationsstudien für das ANOVA–Modell von Einot, Gabriel (1975) und Ramsey (1978) zeigen jedoch, daß dieser Abschlußtest nur unwesentlich besser als der entsprechende Test auf Basis der Homogenitätshypothesen ist. Ein Grund dafür ist, daß nur die Niveaus für die Homogenitätshypothesen geändert werden, die Niveaus für Partitionshypothesen, die keine Homogenitätshypothesen sind, jedoch unverändert bleiben. Für $k \geq 5$ wird zum Beispiel die Hypothese $H_0^{12} \cap H_0^{34} = H_0^{\{1,2\},\{3,4\},\{5\},\ldots,\{k\}}$ nur

zum lokalen Niveau $1-(1-\alpha)^{2/k}(1-\alpha)^{2/k} = 1-(1-\alpha)^{4/k} < \alpha$ geprüft. Das bedeutet aber, daß das multiple Niveau α für $k \geq 5$ nie voll ausgeschöpft wird. Ein weiterer Nachteil dieses Ansatzes ist, daß für eine Partitionshypothese im allgemeinen mehrere verschiedene kritische Werte benötigt werden. Damit werden verschiedene Paarhypothesen $H_0^{ij} \supseteq H_0^{(P_1,\ldots,P_1)}$ im allgemeinen unterschiedlich gewichtet. Da keine Paarhypothese besonders ausgezeichnet ist, erscheint es naheliegender, alle Paarhypothesen $H_0^{ij} \supseteq H_0^{(P_1,\ldots,P_1)}$ gleich stark zu gewichten und für die Partitionshypothesen $H_0^{(P_1,\ldots,P_1)} \in \overline{\mathcal{H}}$ die Niveau–α–Tests

$$\varphi_{(P_1,\ldots,P_1)}(X) := \begin{cases} 1 \\ 0 \end{cases}, \quad T_{(P_1,\ldots,P_1)}(X) \begin{matrix} > \\ \leq \end{matrix} c_\alpha(p_1,\ldots,p_1) \tag{4.8}$$

mit

$$T_{(P_1,\ldots,P_1)}(X) := \max_{1\leq i\leq l} T_{P_i}(X) := \max_{1\leq i\leq l} \max_{r,s\in P_i} |X_r - X_s| \tag{4.9}$$

und

$$c_\alpha(p_1,\ldots,p_1) := \inf\{c : P(\max_{1\leq i\leq l} R_{p_i} \leq c) \geq 1-\alpha\} \tag{4.10}$$

mit $R_1 \equiv 0$ zu verwenden. Da der resultierende Test $\varphi = (\varphi_\pi : \pi \in \Pi_k)$ ein Test zum lokalen Niveau α ist, folgt mit Satz 1.2, daß der durch

$$\overline{\varphi}_\pi := \begin{cases} 1 \ , & \forall\, H_0^{\pi'} \subseteq H_0^{\pi} : \varphi_{\pi'} = 1 \\ 0 \ , & \text{sonst} \end{cases} \tag{4.11}$$

definierte Abschlußtest $\overline{\varphi} := (\overline{\varphi}_\pi : \pi \in \Pi_k)$ ein Test zum multiplen Niveau α für $\overline{H}$ ist. Für das ursprüngliche Testproblem (4.2), also alle Paarvergleiche, ist der entsprechende Test

$$\overline{\varphi} := (\overline{\varphi}_{ij} : 1 \leq i < j \leq k) \tag{4.12}$$

ebenfalls ein Test zum multiplen Niveau α und gleichmäßig besser als der in Abschnitt 3.5 vorgestellte maximax–optimale Test. Den Vorteil, den dieser Ansatz gegenüber dem Vorschlag von Peritz liefert, halten wir in einem Satz fest:

Satz 4.1. Der durch (4.10) definierte Abschlußtest $\bar{\varphi}$ ist lokal konsonant und schöpft, falls F stetig ist, das multiple Niveau α voll aus, d.h.,

$$(4.13) \qquad \forall\, \tau \in \Pi_k : \sup_{\mu \in H_0^\tau} P_\mu(\bar{\varphi}_\tau = 1) = \alpha .$$

Zur Berechnung der kritischen Werte führen wir eine neue Verteilung ein. Im Spezialfall, daß $H_0^{(P_1,\ldots,P_1)} = H_0^{P_j}$ für ein $j \in \{1,\ldots,1\}$ gilt, $H_0^{(P_1,\ldots,P_1)}$ also eine Homogenitätshypothese ist, berechnet sich der entsprechende kritische Wert aus der Verteilungsfunktion der Spannweite mit Parameter $p_j = |P_j|$. Im allgemeinen erhält man für die Partitionshypothesen Maxima von Spannweitenstatistiken. Die zugehörige Verteilung nennen wir Maximum–Range Verteilung mit Parametern $(p_1,\ldots,p_1)$. Mit Hilfe der Verteilungsfunktion der Spannweite erhält man für die Verteilungsfunktion $G_{(p_1,\ldots,p_1)}$ von $T_{(P_1,\ldots,P_1)}$ unter Beachtung der Unabhängigkeit von $T_{P_1},\ldots,T_{P_1}$ für $\mu \in H_0^{(P_1,\ldots,P_1)}$

$$(4.14) \qquad \forall\, z \geq 0 : G_{(p_1,\ldots,p_1)}(z) := P\Big(\max_{1\leq i\leq 1} T_{P_i}(X) \leq z\Big) = \prod_{i=1}^{1} G_{p_i}(z)$$

mit $G_1(z) \equiv 1$, wobei $G_{(p_1,\ldots,p_1)}(z)$ wie $G_{p_i}(z)$ unabhängig von $\mu \in H_0^{(P_1,\ldots,P_1)}$ ist.

Die kritischen Werte der Maximum–Range Verteilung stehen in gewissen Beziehungen zueinander, die wir festhalten wollen:

$$(4.15) \qquad c_\alpha(p_1,\ldots,p_1) \text{ ist unabhängig von der Reihenfolge der } p_i ,$$

(4.16) $\forall\, 1 \le i < j \le l:\ c_\alpha(p_1,\dots,p_i,\dots,p_j,\dots,p_l)$

$$\le c_\alpha(p_1,\dots,p_i+p_j,\dots,p_{j-1},p_{j+1},\dots,p_l)\ ,$$

(4.17) $\forall\, H_0^{(q_1,\dots,q_m)} \subseteq H_0^{(p_1,\dots,p_l)}:\ c_\alpha(p_1,\dots,p_l) \le c_\alpha(q_1,\dots,q_m)\ .$

Die Relationen (4.16) und (4.17) sagen im wesentlichen aus, daß sich durch Zusammenfassen von gewissen p_i bzw. P_i die entsprechenden kritischen Werte vergrößern. Diese Eigenschaften folgen unmittelbar aus der Beziehung

(4.18) $T_{(P_1,P_2)}(X) \le T_{P_1+P_2}(X)$ bzw. $\{T_{(P_1,P_2)}(X) \le c\} \supseteq \{T_{P_1+P_2}(X) \le c\}\ .$

Wegen (4.9) sind die kritischen Werte $c_\alpha(p_1,\dots,p_l)$ schon durch die Angabe aller p_i mit $p_i \ge 2$ eindeutig bestimmt. In praktischen Beispielen wird man also überflüssige "Einsen" im Parameter $(p_1,\dots,p_l)$ weglassen.

Eine neben der Anzahl aller Partitionshypothesen (vgl. Abschnitt 3.2) wichtige Größe im Hinblick auf die praktische Durchführung der Testprozedur ist die Anzahl der verschiedenen kritischen Werte bei fest vorgegebenem k. Diese Zahl entspricht der Anzahl aller Dekompositionen von k in natürliche Zahlen ohne Berücksichtigung der Reihenfolge und ohne die Dekomposition $(1,\dots,1)$ und läßt sich wie folgt rekursiv berechnen (vgl. Abramowitz, Stegun (1968), p. 825):

(4.19) $g(k) = h(k) - 1\ ,\ k = 2,3,\dots$

$$\text{mit } h(k) = \sum_{1 \le \frac{3\ell^2 \pm \ell}{2} \le k} (-1)^{\ell-1}\, h\left(k - \frac{3\ell^2 \pm \ell}{2}\right)\ ,\ k = 1,2,\dots$$

und $h(0) = 1$.

Asymptotisch gilt : $g(k) \sim \dfrac{1}{4k\sqrt{3}} \exp\{\pi\sqrt{2k/3}\}$.

Mit (4.19) erhält man z.B. folgende Werte:

Tabelle 4.1:

k	g(k)
3	2
4	4
5	6
6	10
10	41
13	100
20	626
50	204 225

Zusammenfassend können wir feststellen, daß die formale Definition der abgeschlossenen Spannweitentests keine Probleme bereitet. Die Struktur der einzelnen Tests ist nicht besonders schwierig, die Verteilungsfunktionen der Teststatistiken sind bekannt, die Anzahl der Hypothesen und der kritischen Werte läßt sich berechnen und es gibt keine Einschränkungen an die Ausgangsverteilung F. Im Fall, daß F stetig ist, schöpft der hier vorgeschlagene Abschlußtest im Gegensatz zu allen anderen bisher bekannten Tests das multiple Niveau α für alle $k \geq 3$ voll aus. Der Wermutstropfen liegt in der großen Anzahl der auftretenden Hypothesen sowie der benötigten Anzahl von kritischen Werten, die eine praktische Durchführung der Testprozedur für etwa $k = 20$ mit mehr als $51 \cdot 10^{12}$ Hypothesen zunächst kaum möglich erscheinen läßt. Wir werden im folgenden Abschnitt jedoch zeigen, wie sich die Anzahl der tatsächlich zu prüfenden Hypothesen drastisch verringern läßt.

4.2. Ordnungseigenschaften der lokalen Tests

Letztendlich sind wir nur an den $\binom{k}{2}$ Paarvergleichen und damit am Test $\overline{\psi} = (\overline{\varphi}_{ij} : 1 \leq i < j \leq k)$ interessiert. Einfache Überlegungen zeigen schon, daß nicht alle Hypothesen aus $\overline{\gamma}$ überprüft werden müssen, um die Komponenten von $\overline{\psi}$ zu bestimmen.

Es bietet sich folgende Vorgehensweise an, die schon von Begun, Gabriel (1981) zur Durchführung der von Peritz (1970) vorgeschlagenen Testprozedur im ANOVA–Modell verwendet wurde: Zunächst bestimmt man eine untere und obere Schranke für $\bar{\psi}$, d.h., man bestimmt Tests $\psi^u := (\psi^u_{ij} : 1 \le i < j \le k)$ und $\psi^o := (\psi^o_{ij} : 1 \le i < j \le k)$ mit

$$(4.20) \qquad \forall\, 1 \le i < j \le k \;:\; \psi^u_{ij} \le \bar{\varphi}_{ij} \le \psi^o_{ij} \,.$$

Als untere Schranke verwenden wir das Tukey–Analogon (vgl. (3.7) und (3.8) in Abschnitt 3.1) und als obere Schranke das Analogon zum Newman–Keuls Test der Varianzanalyse (vgl. (3.19) in Abschnitt 3.2). Es wird sich im folgenden als hilfreich erweisen, alle Tests auf der Orderstatistik der X_i aufzubauen und die Hypothesen entsprechend zu indizieren. Unter Verwendung von Kleinbuchstaben für die entsprechenden Statistiken gilt folgende Ordnungseigenschaft der t_{ij}:

$$(4.21) \qquad \forall\, r,s,i,j \in I \text{ mit } 1 \le r \le i < j \le s \le k : t_{rs} \ge t_{ij} \,.$$

Damit erhält man als Newman–Keuls–Analogon

$$(4.22) \qquad \forall\, 1 \le i < j \le k : \psi^o_{ij} := \begin{cases} 1 \;, & \forall\; 1 \le r \le i < j \le s \le k : t_{rs} > c_\alpha(s-r+1) \\ 0 \;, & \text{sonst} \end{cases}$$

und als Tukey–Analogon

$$(4.23) \qquad \forall\, 1 \le i < j \le k : \psi^u_{ij} := \begin{cases} 1 \\ 0 \end{cases} , \; t_{ij} \begin{matrix} > \\ \le \end{matrix} c_\alpha(k) \,.$$

Nach Konstruktion ist ψ^u in jedem Fall ein Test zum multiplen Niveau α (vgl. (3.8) in Abschnitt 3.1), während ψ^o für $k \ge 4$ im allgemeinen das multiple Niveau α nicht einhält. Für $k = 3$ entspricht das Newman–Keuls–Analogon (4.22) schon dem Abschlußtest $\bar{\varphi}$, so daß man sich im folgenden auf $k \ge 4$ beschränken kann.

Eine Hypothese H_0^{uv} mit $\varphi_{uv}^u = 0$ und $\varphi_{uv}^0 = 1$ wird als strittig bezeichnet (vgl. Begun, Gabriel (1981)). Das Problem besteht nun darin, alle strittigen Paarhypothesen auf Annahme oder Ablehnung zu prüfen. Dazu sind formal alle Hypothesen $H_0^\pi \in \mathfrak{H}$ zu prüfen, für die eine strittige Hypothese H_0^{uv} mit $H_0^\pi \subseteq H_0^{uv}$ existiert. Diese Menge läßt sich unter Ausnutzung der Struktur der Spannweitentests leicht einschränken. Ist nämlich $H_0^{(p_1,\ldots,p_l)} \subseteq H_0^{uv}$ und $t_{uv} > c_\alpha(p_1,\ldots,p_l)$, dann gilt $\varphi_{(p_1,\ldots,p_l)} = 1$ und mit (4.17) für alle $(q_1,\ldots,q_m)$ mit $H_0^{uv} \supseteq H_0^{(q_1,\ldots,q_m)} \supseteq H_0^{(p_1,\ldots,p_l)}$ auch $\varphi_{(q_1,\ldots,q_m)} = 1$. Das heißt aber, daß nur noch Hypothesen $H_0^{(q_1,\ldots,q_m)} \subseteq H_0^{uv}$ mit $t_{uv} \leq c_\alpha(q_1,\ldots,q_m)$ zu prüfen sind.

Für jede strittige Hypothese $H_0^{uv} \in \mathfrak{H}$ mit $u < v$ definieren wir

$$(4.24)\quad Q(u,v) := \{\pi = (p_1,\ldots,p_l) \in \Pi_k : H_0^{uv} \supseteq H_0^\pi \text{ und } t_{uv} \leq c_\alpha(p_1,\ldots,p_l)\}\ .$$

Aus den obigen Überlegungen folgen unmittelbar die Aussagen

$$(4.25)\qquad \forall\, \pi \in Q(u,v) : \varphi_\pi = 1 \Rightarrow \bar{\varphi}_{uv} = 1$$

und

$$(4.26)\qquad \exists\, \pi \in Q(u,v) : \varphi_\pi = 0 \Rightarrow \bar{\varphi}_{uv} = 0\ .$$

Um $\bar{\varphi}_{uv}$ zu bestimmen, könnte man also alle H_0^π mit $\pi \in Q(u,v)$ abarbeiten, bis entweder alle Hypothesen abgelehnt worden sind oder aber ein $\pi \in Q(u,v)$ mit $\varphi_\pi = 0$ gefunden wird. Dieses Verfahren wird für größere k allerdings am Umfang der Mengen $Q(u,v)$ scheitern. Unter Ausnutzung der Implikationsstrukturen in $\mathfrak{H}$ und der Ordnungseigenschaft (4.21) der t_{ij} läßt sich jedoch eine Menge $Q^*(u,v) \subseteq Q(u,v)$ konstruieren, so daß nur durch Überprüfung aller H_0^π mit $\pi \in Q^*(u,v)$ entweder $\bar{\varphi}_{uv} = 0$

oder $\overline{\varphi}_{uv} = 1$ folgt. Dabei hat die Menge $\mathfrak{Q}^*(u,v)$ im allgemeinen wesentlich weniger Elemente als $\mathfrak{Q}(u,v)$, so daß eine Durchführung des Abschlußtests auch für größere k möglich wird.

Für $u,v \in I$, $u < v$ sei $(P_1,\ldots,P_l) \in \mathfrak{Q}(u,v)$ mit $\{u,v\} \subseteq P_r$ für ein $r \in \{1,\ldots,l\}$ und $P_r^* := P_r \setminus \{u,v\}$. Mit $\min P := \min_{w \in P} w$ und $\max P := \max_{w \in P} w$ sei $(P_1,\ldots,P_l)$ genau dann ein Element von $\mathfrak{Q}^*(u,v)$, wenn folgende Bedingungen erfüllt sind:

(4.27) $\quad \forall\, s,t \in \{1,\ldots,l\} \setminus \{r\}\ ,\ s < t : \max P_s < \min P_t\ ,$

(4.28) $\quad$ Falls $P_r^* \neq \emptyset \Longrightarrow$
$\forall\, s \in \{1,\ldots,l\} \setminus \{r\} : (\max P_s < \min P_r^*) \vee (\max P_r^* < \min P_s)\ ,$

(4.29) $\quad \forall\, s \in \{1,\ldots,l\} \setminus \{r\} : (\min P_s < \min P_r) \vee (\max P_s > \max P_r)\ ,$

(4.30) $\quad$ Falls $u + 2 \leq v$ und $P_r^* \neq \emptyset \implies (\max P_r^* \geq u - 1) \wedge (\min P_r^* \leq v + 1)\ ,$

(4.31) $\quad$ Falls $u + 1 = v$ und $P_r^* \neq \emptyset \implies (\max P_r^* \geq u - 2) \wedge (\min P_r^* \leq v + 2)\ .$

Bemerkung 4.1. Da für die P_s generell $\min P_s < \min P_{s+1}$ vorausgesetzt war, bedeutet (4.27), daß sich die P_s (ausgenommen P_r) nicht überlappen und in aufsteigender Folge der Indizes geordnet sind. Nach (4.28) folgt dann, daß sich P_r^* in die Reihe der P_s, $s \neq r$, einordnet.

Die Bedingung (4.29) schließlich bedeutet, daß keines der P_s ganz zwischen $\min P_r$ und $\max P_r$ liegt. Daraus folgt zusammen mit (4.27) und (4.28), daß P_r nur mit P_{r-1} und/oder P_{r+1} überlappt und daß P_r^* zwischen P_{r-1} und P_{r+1} liegt. Also können höchstens u und v in den Bereichen von P_{r-1} und P_{r+1} liegen. Liegen u und v beide rechts oder links von P_r^*, so gibt es wegen (4.30) im Fall $u + 2 \leq v$ kein Element, und wegen (4.31) im Fall $u + 1 = v$ höchstens ein Element zwischen $\{u,v\}$ und P_r^*. □

Satz 4.2. Seien $u,v \in I$, $u < v$ mit $\bar{\varphi}_{uv} = 0$. Dann gilt:
$\exists\, (P_1,\ldots,P_l) \in \mathfrak{Q}^*(u,v) : \varphi_{(P_1,\ldots,P_l)} = 0.$

Beweis: Ist $\bar{\varphi}_{uv} = 0$, so folgt aus der Konstruktion von $\bar{\varphi}$ die Existenz einer Partition $(M_1,\ldots,M_m) \in \mathfrak{Q}(u,v)$ mit $\varphi_{(M_1,\ldots,M_m)} = 0$. Diese Partition läßt sich in eine neue Partition $(P_1,\ldots,P_l) \in \mathfrak{Q}^*(u,v)$ überführen, so daß

(4.32) $\quad t_{(P_1,\ldots,P_l)} \leq t_{(M_1,\ldots,M_m)}$

und

(4.33) $\quad c_\alpha(p_1,\ldots,p_l) \geq c_\alpha(m_1,\ldots,m_m)$.

Wegen $\varphi_{(M_1,\ldots,M_m)} = 0$ folgt damit unmittelbar $\varphi_{(P_1,\ldots,P_l)} = 0$.

Zur Konstruktion der Partition $(P_1,\ldots,P_l) \in \mathfrak{Q}^*(u,v)$ mit (4.32) und (4.33) definieren wir folgende Durchmesser für die Mengen P_i und die Partitionen $\pi \in \Pi_k$:

(4.34) $\quad \forall\, \emptyset \neq P \subseteq I : d(P) := \max P - \min P + 1$.

und

(4.35) $\quad \forall\, (P_1,\ldots,P_l) \in \Pi_k : d(P_1,\ldots,P_l) := \sum_{i=1}^{l} d(P_i)$.

Offensichtlich gilt

(4.36) $\quad \forall\, (P_1,\ldots,P_l) \in \Pi_k : d(P_1,\ldots,P_l) \geq k$ mit "=" genau dann, wenn
$\forall\, s,t \in \{1,\ldots,l\}$, $s < t : \max P_s < \min P_t$.

Sei nun $(M_1,\ldots,M_m) \in \mathfrak{Q}(u,v)$ mit $\varphi_{(M_1,\ldots,M_m)} = 0$ und $(M_1,\ldots,M_m) \notin \mathfrak{Q}^*(u,v)$.
Dann werden wir zeigen:

(4.37) $\exists\ (N_1,\ldots,N_n) \in \mathfrak{Q}(u,v)$ mit

(a) $d(N_1,\ldots,N_n) \leq d(M_1,\ldots,M_m) - 1$

(b) $t_{(N_1,\ldots,N_n)} \leq t_{(M_1,\ldots,M_m)}$

(c) $c_a(n_1,\ldots,n_n) \geq c_a(m_1,\ldots,m_m)$.

Hieraus folgt induktiv die Existenz eines $(P_1,\ldots,P_l) \in \mathfrak{Q}^*(u,v)$ mit $\varphi_{(P_1,\ldots,P_l)} = 0$, also die Behauptung des Satzes, da spätestens bei $d(N_1,\ldots,N_n) = k$ wegen (4.36) diese Eigenschaft erreicht wird.

Um (4.37) zu zeigen, sind einige Fallunterscheidungen notwendig.
Ist $(M_1,\ldots,M_m) \notin \mathfrak{Q}^*(u,v)$, so ist mindestens eine der Bedingungen (4.27) – (4.30) bzw. (4.27) – (4.29) + (4.31) verletzt. Es genügt, die folgenden 4 Fälle zu betrachten:

(α) (4.27) ist verletzt,

(β) (4.28) ist verletzt,

(γ) (4.27) + (4.28) sind erfüllt, (4.29) ist verletzt,

(δ) (4.27) – (4.29) sind erfüllt, (4.30) bzw. (4.31) ist verletzt.

In allen Fällen lassen sich aus $(M_1,\ldots,M_m)$ Mengen $N_1,\ldots,N_n$ konstruieren, die nach eventueller Indextransformation von $1,\ldots,n$ eine Partition $(N_1,\ldots,N_n) \in \mathfrak{Q}(u,v)$ mit (a), (b), (c) aus (4.37) ergeben. Dabei beschränken wir uns auf die Angabe der Mengen $N_1,\ldots,N_n$ und überlassen den Nachweis von (4.37) dem Leser.
Im folgenden sei $\{u,v\} \subseteq M_r$ und $M_r^* := M_r \setminus \{u,v\}$.

(α) Ist (4.27) verletzt, so existieren $s,t \in \{1,\ldots,m\} \setminus \{r\}$, $s < t$: $\max M_s > \min M_t$.
Seien $M_s = \{w_1,\ldots,w_\alpha\}$, $M_t = \{w_{\alpha+1},\ldots,w_{\alpha+\beta}\}$ mit $w_1 < \cdots < w_\alpha$ und $w_{\alpha+1} < \cdots < w_{\alpha+\beta}$. Dann ist $w_1 < w_{\alpha+1}$ und nach Vor. $w_{\alpha+1} < w_\alpha$. Ordnet man die Elemente von $M_s + M_t$ der Größe nach und bezeichnet diese mit

$w_{(1)} < \ldots < w_{(\alpha+\beta)}$, so setzt man

$$N_i := \begin{cases} \{w_{(1)},\ldots,w_{(\alpha)}\} & , i = s \\ \{w_{(\alpha+1)},\ldots,w_{(\alpha+\beta)}\} & , i = t \\ M_i & , \text{sonst} . \end{cases}$$

(β) Ist (4.28) verletzt, so existiert ein $s \in \{1,\ldots,m\} \setminus \{r\}$ mit

$(\max M_s > \min M_r^*) \wedge (\max M_r^* > \min M_s)$.

Aus Symmetriegründen kann man o.B.d.A. $s < r$ annehmen.

Sei $M_s := \{w_1,\ldots,w_\alpha\}$ und $M_r^* = \{w_{\alpha+1},\ldots,w_{\alpha+\beta}\}$ mit $w_1 < \cdots < w_\alpha$ und $w_{\alpha+1} < \cdots < w_{\alpha+\beta}$. Dann ist $w_1 < w_{\alpha+1}$ und nach Vor. $w_{\alpha+1} < w_\alpha$. Man ordnet $M_s + M_r^*$ wie in (α) und setzt

$$N_i := \begin{cases} \{w_{(1)},\ldots,w_{(\alpha)}\} & , i = s \\ \{w_{(\alpha+1)},\ldots,w_{(\alpha+\beta)}\} & , i = r \\ M_i & , \text{sonst} . \end{cases}$$

(γ) Ist (4.27) und (4.28) erfüllt und (4.29) verletzt, so existiert ein $t \geq 1$ so daß

$\{r+1,\ldots,r+t\} = \{s \in \{1,\ldots,m\}\setminus\{r\} : (\min M_s > \min M_r) \wedge (\max M_s < \max M_r)$.

Wähle

$$N_i := \begin{cases} M_i & , i < r \\ M_r + \sum_{j=1}^{t} M_{r+j} & , i = r \\ M_{t+i} & , r < i \leq l-t . \end{cases}$$

(δ) Seien (4.27) – (4.29) erfüllt.

(i) Ist $u + 2 \leq v$ und (4.30) verletzt, so gilt: $(\max M_r^* \leq u-2) \vee (\min M_r^* \geq v + 2)$.

Sei $M_r^* := \{w_1,\cdots,w_\alpha\}$ mit $w_1 < \cdots < w_\alpha$. Aus Symmetriegründen nehmen wir o.B.d.A. $\max M_r^* \leq u - 2$ an. Wegen (4.27) – (4.29) existiert ein $s \in \{1,\cdots,m\}, s > r$, so daß $\{u-1,u+1,v+1\} \subseteq M_s := \{w_{\alpha+1},\ldots,w_{\alpha+\beta}\}$.

Setze

$$N_i := \begin{cases} M_r^* + \{w_{\alpha+1}, w_{\alpha+2}\} & , i = r \\ M_s \setminus \{w_{\alpha+1}, w_{\alpha+2}\} + \{u,v\} & , i = s \\ M_i & , \text{sonst} . \end{cases}$$

(ii) Ist $u+1 = v$ und (4.31) verletzt, so gilt $(\max M_r^* \leq u-3) \vee (\min M_r^* \geq v+3)$. Sei o.B.d.A. $\max M_r^* \leq u-3$, und M_r^* wie in (i). Wegen (4.27) – (4.29) existiert ein $s \in \{1,\ldots,m\}$, $s > r$, so daß $\{u-2, u-1, v+1\} \subseteq M_s := \{w_{\alpha+1},\ldots,w_{\alpha+\beta}\}$. Setze

$$N_i := \begin{cases} M_r^* + \{w_{\alpha+1}, w_{\alpha+2}\} & , i = r \\ M_s \setminus \{w_{\alpha+1}, w_{\alpha+2}\} + \{u,v\} & , i = s \\ M_i & , \text{sonst} . \end{cases}$$

Insgesamt folgt die Behauptung des Satzes. □

Bemerkung 4.2. Hat man eine Partition $(P_1,\ldots,P_l) \in \mathfrak{P}^*(u,v)$ mit $\varphi_{(P_1,\ldots,P_l)} = 0$ gefunden, so folgt natürlich sofort

$$(4.38) \qquad \forall\, H_0^{ij} \supseteq H_0^{(P_1,\ldots,P_l)} : \bar{\varphi}_{ij} = 0 . \quad \square$$

Die Ordnungseigenschaft (4.21) der t_{ij} legt die Frage nahe, ob man nicht aus $\bar{\varphi}_{uv} = 0$ sogar $\bar{\varphi}_{ij} = 0$ für alle $u \leq i < j \leq v$ folgern kann. Ein einfaches, wenn auch konstruiertes Beispiel widerlegt diese Annahme.

Beispiel 4.1. Sei $f(x) = \eta\, I_{(0,1)}(x) + (1-\eta)\, I_{(3,4)}(x)$, $\eta \in (0,1)$, Lebesgue–Dichte von P. Für $0 \leq z \leq 1$ erhält man mit Hilfe von (2.4) für die Verteilungsfunktion der Spannweite

$$G_k(z) = k\, z^{k-1} \left[\eta^k + (1-\eta)^k\right] \left[1 - (1 - \tfrac{1}{k})\, z\right] .$$

Wählt man $\alpha = 0.05$ und η so, daß $G_4(0.98) = 0.95$, so erhält man $\eta \approx 0.9878$ und im Fall von 4 Mittelwerten die kritischen Werte $c_\alpha(4) = 0{,}98$, $c_\alpha(3) \approx 0{,}9288$, $c_\alpha(2{,}2) = 0{,}9637$ und $c_\alpha(2) = 0.8369$. Die Realisationen der Beobachtungen seien gegeben durch $x_1 = 0.005$, $x_2 = 0.04$, $x_3 = 0.96$ und $x_4 = 0.995$. Für $\bar{\varphi}$ erhält man $\bar{\varphi}_{14} = \bar{\varphi}_{23} = 1$ und $\bar{\varphi}_{12} = \bar{\varphi}_{13} = \bar{\varphi}_{24} = \bar{\varphi}_{34} = 0$. Obwohl also $t_{13} = t_{24} > t_{23}$ ist, wird H_0^{23} abgelehnt, während H_0^{13} und H_0^{24} nicht abgelehnt werden können. Die Erklärung für dieses doch wohl überraschende Entscheidungsmuster liegt in der Eigenschaft, daß $c_\alpha(3) < c_\alpha(2{,}2)$ ist. □

Im folgenden werden wir zeigen, daß Entscheidungen wie in Beispiel 4.1 nur für spezielle Indizes $i < j$ zwischen u und v möglich und bei einigen Verteilungen generell ausgeschlossen sind.

Satz 4.3. Seien $u, v \in I$, $u + 3 \leq v$ mit $\bar{\varphi}_{uv} = 0$. Dann gilt:

$$(4.39) \qquad \forall (i,j) \in K_{uv} := \{(r,s) : u \leq r < r+1 < s \leq v\} : \bar{\varphi}_{ij} = 0 .$$

Beweis: Wir zeigen zunächst, daß aus $\bar{\varphi}_{uv} = 0$ auch $\bar{\varphi}_{u,v-1} = 0$ folgt. Nach Satz 4.2 existiert ein $(P_1, \ldots, P_l) \in \mathfrak{P}^*(u,v)$ mit $\{u,v\} \subseteq P_r$, $r \in \{1, \ldots, l\}$ und $\varphi_{(P_1,\ldots,P_l)} = 0$. Ist $v-1 \in P_r$, so folgt $\bar{\varphi}_{u,v-1} = 0$. Ist $v-1 \notin P_r$, so gilt wegen (4.27) – (4.30) und $u+3 \leq v$ entweder $v-1 \in P_{r-1} \supseteq \{u-1, u+1, v-1\}$ mit $u+1 < v-1$ oder $v-1 \in P_{r+1} \supseteq \{v-1, v+1\}$. Im ersten Fall setzt man

$$P_i' := \begin{cases} P_{r-1} \setminus \{u+1\} + \{u\} & , i = r-1 \\ P_r \setminus \{u\} + \{u+1\} & , i = r \\ P_i & , \text{sonst} , \end{cases}$$

im zweiten Fall

$$P_i' := \begin{cases} P_{r+1} \setminus \{v-1\} + \{v\} & , i = r-1 \\ P_r \setminus \{v\} + \{v-1\} & , i = r \\ P_i & , \text{sonst} . \end{cases}$$

In beiden Fällen gilt $\varphi_{(P_1', \ldots, P_l')} = 0$ und damit $\bar{\varphi}_{u,v-1} = 0$. Ist $(i,j) \in K_{uv}$ und $j < v-1$, so erhält man induktiv $\bar{\varphi}_{u,v-2} = 0 , \ldots, \bar{\varphi}_{uj} = 0$. Ist $i > u$, so folgt aus Symmetriegründen analog $\bar{\varphi}_{u+1,j} = 0 , \ldots, \bar{\varphi}_{ij} = 0$. □

Bemerkung 4.3. Ist $u,v \in I$, $u + 3 \leq v$ mit $\bar{\varphi}_{uv} = 0$, so ist $\bar{\varphi}_{ij} = 1$ für $u \leq i < j \leq v$ nach Satz 4.3 höchstens dann möglich, wenn $j = i + 1$, und daß dieser Fall vorkommt, haben wir in Beispiel 4.1 gezeigt. □

Falls die kritischen Werte gewisse Monotonieeigenschaften erfüllen, so daß Fälle wie in Beispiel 4.1 nicht auftreten, so werden wir zeigen, daß aus $\bar{\varphi}_{uv} = 0$ stets $\bar{\varphi}_{ij} = 0$ für alle $u \leq i < j \leq v$ folgt. Eine mögliche Bedingung an die kritischen Werte ist, daß zu gegebenem $\alpha \in (0,1)$ gilt

$$(4.40)\ \forall\, (p_1, \ldots, p_i, \ldots, p_j, \ldots, p_l) \in N_k , l \geq 2, 2 \leq p_i \leq p_j :$$
$$c_\alpha(p_1, \ldots, p_i, \ldots, p_j, \ldots, p_l) \leq c_\alpha(p_1, \ldots, p_i - 1, \ldots, p_j + 1, \ldots, p_l) .$$

Diese Bedingung läßt sich für alle $\alpha \in (0,1)$ aus der logarithmischen Konkavität der Verteilungsfunktion der Spannweite $G_k(z)$ in k folgern:

Lemma 4.1. Ist $G_k(z)$ für alle $z \geq 0$ log–konkav in k, so folgt (4.40).

Beweis: Aus

$$\forall\, r,s \in \mathbb{N}\,,\ 2 \le r \le s:\ \forall\, z \ge 0 : G_r(z)\, G_s(z) \ge G_{r-1}(z)\, G_{s+1}(z)\ .$$

folgt für die Verteilungsfunktion $G_{(p_1,\ldots,p_l)}(z) = \prod_{i=1}^{l} G_{p_i}(z)$ der Maximum–Range-Verteilung

$$(4.41)\quad \forall\, (p_1,\ldots,p_i,\ldots,p_j,\ldots,p_l) \in \mathbb{N}_k\,,\ l \ge 2,\ 2 \le p_i \le p_j :$$
$$G_{(p_1,\ldots,p_i,\ldots,p_j,\ldots,p_l)}(z) \ge G_{(p_1,\ldots,p_i-1,\ldots,p_j+1,\ldots,p_l)}(z)\ ,$$

also auch (4.40). □

Fordert man in Satz 4.2 zusätzlich (4.40), so erhält man folgende verschärfte Version:

Satz 4.4. Sei (4.40) erfüllt, $u,v \in I$, $u < v$ mit $\bar{\varphi}_{uv} = 0$.
Dann gilt: $\exists\, (P_1,\ldots,P_l) \in \mathfrak{Q}^*(u,v)$ mit $\varphi_{(P_1,\ldots,P_l)} = 0$ und

$$(4.42)\qquad \forall\, s \in \{1,\ldots,l\} \setminus \{r\} : (|P_r + P_s^*| < |P_s|) \vee (P_s^* = \emptyset)$$
$$\text{mit } \{u,v\} \subseteq P_r,\ P_s^* = \{w \in P_s : \min P_r < w < \max P_r\}\ .$$

Beweis: Nach Satz 4.2 existiert ein $(P_1,\ldots,P_l) \in \mathfrak{Q}^*(u,v)$ mit $\varphi_{(P_1,\ldots,P_l)} = 0$. Ist (4.42) für $(P_1,\ldots,P_l)$ nicht erfüllt, so existiert ein $s \in \{1,\ldots,l\} \setminus \{r\}$ mit $|P_r + P_s^*| \ge |P_s|$ und $P_s^* \neq \emptyset$. Aus der Definition von $\mathfrak{Q}^*(u,v)$ folgt $s \in \{r-1, r+1\}$. Sei o.B.d.A. $s = r-1$. Setze

$$P_i' := \begin{cases} P_{r-1} \setminus P_{r-1}^* & , i = r-1 \\ P_r + P_{r-1}^* & , i = r \\ P_i & , \text{sonst}\ . \end{cases}$$

Dann folgt wegen $\max\{p_{r-1},p_r\} \le \max\{p'_{r-1},p'_r\}$ aus (4.40) $c_a(p_1,\dots,p_l) \le c_a(p'_1,\dots,p'_l)$, und wegen $t_{(p'_1,\dots,p'_l)} \le t_{(p_1,\dots,p_l)}$ erhält man $\varphi_{(p'_1,\dots,p'_l)} = 0$. Erfüllt $(p'_1,\dots,p'_l)$ die Bedingung (4.42), so folgt die Behauptung. Anderenfalls ist $P'_{r+1} \neq \emptyset$ und $|P'_r + P'^{*}_{r+1}| \ge |P'_{r+1}|$. Setzt man

$$P''_i := \begin{cases} P'_r + P'^{*}_{r+1} & , i = r \\ P'_{r+1} \setminus P'^{*}_{r+1} & , i = r+1 \\ P'_i & , \text{sonst} , \end{cases}$$

so folgt wie im ersten Schritt $\varphi_{(p''_1,\dots,p''_l)} = 0$. Insbesondere erfüllt $(P''_1,\dots,P''_l)$ die Bedingung (4.42). □

Satz 4.5. Sei (4.40) erfüllt, $u,v \in I$, $u+2 \le v$ mit $\bar{\varphi}_{uv} = 0$. Dann gilt:

(4.43) $\quad \forall\, u \le i < j \le v : \bar{\varphi}_{ij} = 0.$

Beweis: Wegen Satz 4.3 genügt es, Satz 4.5 für $u+2 = v$ zu zeigen. Nach Satz 4.4 existiert ein $(P_1,\dots,P_l)$ mit (4.42) und $\varphi_{(p_1,\dots,p_l)} = 0$. Sei wieder $\{u,v\} \subseteq P_r$ für ein $r \in \{1,\dots,l\}$. Ist $u+1 \in P_r$, so folgt die Behauptung. Ist $u+1 \notin P_r$, so existiert ein $s \in \{r-1,r+1\}$ mit $u+1 \in P_s$. Sei o.B.d.A. $s = r-1$. Dann ist $P_{r-1} = \{w_1, w_1+1, \dots, u-1, u+1\}$ mit $w_1 \le u-1$ und $P_r = \{u, u+2, \dots, u+w_2\}$ mit $w_2 \ge 2$.
Wegen $P^{*}_{r-1} := \{w \in P_{r-1} : \min P_r < w < \max P_r\} = \{u+1\} \neq \emptyset$ folgt aus (4.42)
$|P_r + P^{*}_{r-1}| = |P_r| + 1 < |P_{r-1}|$. Setzt man

$$P'_i := \begin{cases} P_{r-1} \setminus \{u+1\} + \{u\} & , i = r-1 \\ P_r \setminus \{u\} + \{u+1\} & , i = r \\ P_i & , \text{sonst} , \end{cases}$$

so folgt wegen $t_{(P_1,\ldots,P_l)} \geq t_{(P'_1,\ldots,P'_l)}$ sofort $\varphi_{(P'_1,\ldots,P'_l)} = 0$, also auch $\bar{\varphi}_{u+1,v} = \bar{\varphi}_{u+1,u+2} = 0$.

Setzt man

$$P''_i := \begin{cases} P_{r-1} + \{u\} & , i = r-1 \\ P_r \setminus \{u\} & , i = r \\ P_i & , \text{sonst} , \end{cases}$$

so erhält man wegen $|P_r| + 1 < |P_{r-1}|$ aus (4.40) und $t_{(P_1,\ldots,P_l)} \geq t_{(P''_1,\ldots,P''_l)}$, daß $\varphi_{(P''_1,\ldots,P''_l)} = 0$, also auch $\bar{\varphi}_{u,u+1} = 0$. □

Eine für die praktische Durchführung der Testprozedur wichtige Folgerung aus den bisherigen Resultaten halten wir im folgenden Korollar fest:

Korollar 4.1. Für $u,v \in I$, $u < v$ sei $J_{uv} := \{(r,s) : 1 \leq r \leq u < v \leq s \leq k\}$.
(a) Falls $u+2 \leq v$ und $\bar{\varphi}_{uv} = 1$, so folgt : $\forall\, (i,j) \in J_{uv} : \bar{\varphi}_{ij} = 1$.
(b) Falls $u+1 = v$, $\bar{\varphi}_{uv} = 1$ und (4.40) erfüllt ist, so folgt: $\forall\, (i,j) \in J_{uv} : \bar{\varphi}_{ij} = 1$.

Mit Hilfe von Korollar 4.1 und einer geschickten Testreihenfolge der strittigen Hypothesen läßt sich die Anzahl der zu überprüfenden Hypothesen noch einmal verringern.
Sei $\mathfrak{Q}_1^{**}(u,v)$ die Menge aller $(P_1,\ldots,P_l) \in \mathfrak{Q}^*(u,v)$, die zusätzlich die Bedingung

(4.44) $\qquad \min P_r = u \;\wedge\; \max P_r = v$

erfüllen, und $\mathfrak{Q}_2^{**}(u,v)$ die Menge aller $(P_1,\ldots,P_l) \in \mathfrak{Q}_1^{**}(u,v)$, für die zusätzlich (4.42) gilt.

Korollar 4.2. Sei $u,v \in I$, $u < v$ und $\bar{\varphi}_{uv} = 0$.

(a) Falls $\forall (i,j) \in J_{uv} \setminus \{u,v\} : \bar{\varphi}_{ij} = 1$, so folgt:

$\exists (P_1,\ldots,P_l) \in Q_1^{**}(u,v) : \varphi_{(P_1,\ldots,P_l)} = 0.$

(b) Ist (4.40) erfüllt und gilt $\forall (i,j) \in J_{uv} \setminus \{u,v\} : \bar{\varphi}_{ij} = 1$, so folgt:

$\exists (P_1,\ldots,P_l) \in Q_2^{**}(u,v) : \varphi_{(P_1,\ldots,P_l)} = 0.$

Im folgenden Abschnitt wird gezeigt, wie sich die bisherigen Ergebnisse für einen praktischen Algorithmus ausnutzen lassen.

4.3. Ein Algorithmus.

Gilt für eine strittige Hypothese H_0^{uv} letztendlich doch $\bar{\varphi}_{uv} = 1$, so spielt die Reihenfolge der Überprüfung aller Partitionen aus $Q_1^{**}(u,v)$ bzw. $Q_2^{**}(u,v)$ praktisch keine Rolle, da alle Partitionen überprüft werden müssen. Ist jedoch $\bar{\varphi}_{uv} = 0$, so wäre es wünschenswert, möglichst schnell eine Partition $(P_1,\ldots,P_l)$ aus $Q_1^{**}(u,v)$ bzw. $Q_2^{**}(u,v)$ mit $\varphi_{(P_1,\ldots,P_l)} = 0$ zu finden, da in diesem Fall die verbleibenden Partitionen nicht weiter überprüft werden müssen. Es liegt daher nahe, vor der Totalüberprüfung aller strittigen Hypothesen eine heuristisch begründete Auswahl von möglichst einfach strukturierten Hypothesen zu treffen und diese auf Annahme zu prüfen. Es bieten sich zum Beispiel alle Hypothesen $H_0^{(P_1,\ldots,P_l)}$ mit $2 \le l \le 4$ und

(4.45) $\quad \forall s,t \in \{1,\ldots,l\}, s < t : \max P_s < \min P_t$

an. Ist $\varphi_{(P_1,\ldots,P_l)} = 0$, so folgt : $\forall H_0^{ij} \supseteq H_0^{(P_1,\ldots,P_l)} : \bar{\varphi}_{ij} = 0$.

Praktische Erfahrungen zeigen, daß sich damit die in (4.22) angegebene obere Schranke ψ^0 häufig deutlich verbessern läßt. Gleichzeitig verringert sich natürlich auch

die Zahl der strittigen Hypothesen. Bei der Wahl einer solchen Methode ist jedoch darauf zu achten, daß das Verhältnis von Rechenaufwand und Reduzierung der strittigen Hypothesen in einem für die Rechenzeit durchschnittlich günstigen Verhältnis bleibt.

Für die weitere Überprüfung der strittigen Hypothesen bietet sich folgende Vorgehensweise an. Zunächst wird die strittige Hypothese mit dem größten t–Wert überprüft. Gibt es mehrere strittige Hypothesen H_0^{uv} mit dem gleichen t–Wert, so wird die Hypothese ausgewählt, für die v–u maximal ist. Anschließend werden alle Hypothesen $H_0^{(P_1,\ldots,P_l)}$ mit $(P_1,\ldots,P_l) \in \mathfrak{Q}_1^{**}(u,v)$ bzw. $\in \mathfrak{Q}_2^{**}(u,v)$ gemäß der Reihenfolge der kritischen Werte abgearbeitet. Dabei seien die kritischen Werte der Größe nach in fallender Reihenfolge geordnet. Findet sich eine Partition $(P_1,\ldots,P_l)$ mit $\varphi_{(P_1,\ldots,P_l)} = 0$, so wird überprüft, ob durch Anwendung von Bemerkung 4.2 oder Satz 2.4 weitere strittige Hypothesen angenommen werden müssen. Ist entschieden, ob $\bar{\varphi}_{uv} = 0$ oder $\bar{\varphi}_{uv} = 1$ gilt, so wird wie oben aus den verbliebenen strittigen Hypothesen wieder diejenige mit dem größten t–Wert ausgewählt und wie oben überprüft. Dieses Verfahren wird so lange fortgesetzt, bis für alle strittigen Hypothesen entweder $\bar{\varphi}_{uv} = 0$ oder $\bar{\varphi}_{uv} = 1$ nachgewiesen ist.

Der gesamte Algorithmus läßt sich grob in 5 Punkte einteilen:

(I) Bestimme mit (4.23) eine untere Schranke $\psi^u := (\psi^u_{12},\ldots,\psi^u_{k-1,k})$ von $\bar{\psi}$.

(II) Bestimme mit (4.22) eine obere Schranke $\psi^o := (\psi^o_{12},\ldots,\psi^o_{k-1,k})$ von $\bar{\psi}$.

(III) Klassifiziere alle Hypothesen mit $\psi^u_{uv} = 0$ und $\psi^o_{uv} = 1$ als strittig.

(IV) Versuche die obere Schranke ψ^o von $\bar{\psi}$ durch ein möglichst einfaches und wenig rechenaufwendiges heuristisches Verfahren zu verbessern, um die Zahl der strittigen Hypothesen zu verringern.

(V) Überprüfe alle strittigen Hypothesen gemäß der Reihenfolge der t–Werte und der kritischen Werte durch Überprüfung der entsprechenden Mengen $\mathfrak{Q}_1^{**}(u,v)$ bzw. $\mathfrak{Q}_2^{**}(u,v)$. Verwende dabei insbesondere Bemerkung 4.2 und Satz 4.3 bzw. Satz 4.4.

Im nächsten Abschnitt werden die bisherigen Ergebnisse auf das ANOVA–Modell übertragen. Anschließend wird die Arbeitsweise des hier vorgestellten Algorithmus an einem Beispiel mit 13 Mittelwerten illustriert.

4.4. Der abgeschlossene Newman–Keuls Spannweitentest im ANOVA–Modell mit gleichen Stichprobenumfängen.

Das ANOVA–Modell mit gleichen Stichprobenumfängen gehört wohl zu den in der Praxis am häufigsten betrachteten Modellen. So wird denn auch für das Problem der paarweisen Vergleiche eine Vielzahl von Lösungen angeboten. Im Vergleich zu den bisher betrachteten Problemen kommt als Erschwernis ein unbekannter Skalenparameter hinzu. Viele in der Praxis betrachteten Modelle lassen sich auf folgende Situation zurückführen:

(4.46) $X_i \sim N(\mu_i, \sigma^2)$, $i=1,\ldots,k$ und $S^2 \sim \frac{\sigma^2}{\nu} \chi^2_\nu$ seien stochastisch unabhängige Zufallsvariable, wobei $\mu_i \in \mathbb{R}$, $i=1,\ldots,k$ und $\sigma^2 > 0$ unbekannte Parameter seien.

Gesucht ist wiederum ein abgeschlossener multipler Spannweitentest für die Paarvergleiche. Standardisiert man alle in Abschnitt 4.1 benötigten Teststatistiken mit Hilfe von S, so gelangt man zu studentisierten Maximum–Range Statistiken, in Spezialfällen zur aus dem Tukey–Test bekannten studentisierten Spannweitenstatistik oder auch zur einfachen t–Statistik. In jedem Fall bleiben alle in Abschnitt 4.2 hergeleiteten Resultate auch in dieser Situation richtig (insbesondere erfüllen die entsprechenden kritischen Werte die Monotoniebedingung (4.40)) und der Algorithmus läßt sich wie in Abschnitt 4.3 beschrieben durchführen. Die einzige Veränderung liegt in der etwas komplizierteren Verteilung der Statistiken und der Berechnung der kritischen Werte. Hervorzuheben

ist, daß auch hier durch die Verwendung der exakten Verteilungen das multiple Niveau voll ausgeschöpft wird. In Analogie zu Abschnitt 4.1 erhält man die Statistiken

$$(4.47) \qquad \forall\, 1 \le i < j \le k \quad : T_{ij}(X) = \frac{|X_i - X_j|}{S}$$

$$(4.48) \qquad \forall\, \emptyset \neq P \subseteq I \quad : T_P(X) := \max_{i,j \in P} T_{ij}(X)$$

$$(4.49) \qquad \forall\, (P_1,\ldots,P_l) \in \Pi_k : T_{(P_1,\ldots,P_l)}(X) := \max_{1 \le i \le l} T_{P_i}(X) \,.$$

Unter den entsprechenden Nullhypothesen genügen $(X_i - X_j)/(\sqrt{2}\, S)$ bzw. T_P einer t–Verteilung mit ν bzw. einer studentisierten Spannweitenverteilung mit $p := |P|$ und ν Freiheitsgraden. Die Verteilung von $T_{(P_1,\ldots,P_l)}$ entspricht im Spezialfall $|P_i| \le 2$, $i=1,\ldots,l$ der bekannten Maximum–Modulus–Verteilung (vgl. z.B. Miller (1966)) mit entsprechenden Freiheitsgraden. Da die Verteilungen unabhängig von $\sigma^2 > 0$ sind, kann bei ihrer Berechnung o.B.d.A. $\sigma^2 = 1$ angenommen werden. Unter Ausnutzung der Unabhängigkeit von $X_1,\ldots,X_k$ und S^2 erhält man für beliebiges $(P_1,\ldots,P_l) \in \Pi_k$ unter $\mu \in H_0^{(P_1,\ldots,P_l)}$

$$(4.50) \qquad P_\mu\,(T_{(P_1,\ldots,P_l)}(X) \le z)$$

$$= P_\mu\Big(\bigcap_{i=1}^{l} \{\max_{r,s \in P_i} |X_r - X_s| \le z \cdot S\}\Big)$$

$$= \int_{(0,\infty)} P_\mu\Big(\bigcap_{i=1}^{l} \{\max_{r,s \in P_i} |X_r - X_s| \le zs\}\Big)\, dP^S(s)$$

$$= \int_{(0,\infty)} \prod_{i=1}^{l} P_\mu\,(\{\max_{r,s \in P_i} |X_r - X_s| \le zs\})\, dP^S(s)$$

$$= \int_{(0,\infty)} \prod_{i=1}^{l} G_{p_i}(zs)\, dP^S(s)$$

$$= \int_0^\infty \prod_{i=1}^{l} G_{p_i}(zs)\, f_\nu(s)\, ds$$

$$=: G_{(p_1,\ldots,p_l;\nu)}(z)$$

mit

$$(4.51) \qquad \forall\, p \in \mathbb{N} : \forall\, z \geq 0 : G_p(z) = p \int_{-\infty}^{\infty} \varphi(x)\, [\Phi(x) - \Phi(x-z)]^{p-1}\, dx$$

$$(4.52) \qquad \varphi(x) = (2\pi)^{-1/2} \exp\{-\frac{x^2}{2}\}$$

$$(4.53) \qquad \Phi(x) = \int_{-\infty}^{x} \varphi(x)\, dx$$

$$(4.54) \qquad f_\nu(s) = \frac{\nu^{\nu/2}}{\Gamma(\frac{\nu}{2})\ 2^{\nu/2-1}}\ s^{\nu-1} \exp\{-\nu\,\frac{s^2}{2}\} \cdot I_{(0,\infty)}(s)\ .$$

$G_{(p_1,\ldots,p_1;\nu)}$ ist offensichtlich unabhängig von $\mu \in H_0^{(p_1,\ldots,p_1)}$.

Die Berechnung und Tabellierung von kritischen Werten für die studentisierte Maximum–Range–Verteilung ist recht aufwendig und wurde vom Autor dieser Arbeit für $\alpha = 0.1$, 0.05 und 0.01 , $k = 2(1)20$ und $\nu = 1(1)20(2)40(5)120$ und $\nu = \infty$ durchgeführt. Da die entsprechenden Tabellen ein ganzes Buch füllen würden, sind diese Werte nur in Diskettenform vom Autor erhältlich. Eine Tabelle für $\alpha = 0.05$, $k = 2(1)10$ und ausgewählte Freiheitsgrade von ν findet man in Royen (1987), wo auf der Basis der studentisierten Maximum–Range–Verteilung eine Verbesserung der von Shaffer (1986) vorgestellten modifizierten Bonferroni–Holm Prozedur vorgestellt wird, die jedoch dem hier vorgestellten Abschlußtest gleichmäßig unterlegen ist. Ein stichprobenartiger Vergleich mit den von Royen tabellierten Werten ergab in wenigen Fällen Unterschiede von $\pm\ 0.001$. Mit den von Harter (1960) tabellierten kritischen Werten der studentisierten Spannweite ergaben sich ebenfalls in einigen Fällen Unterschiede von $\pm\ 0.001$.

Die Berechnung wurde wie folgt durchgeführt. Um $G_{(p_1,\ldots,p_1;\nu)}(z)$ zu berechnen, wurden zunächst die inneren Integrale (4.51) für festes zs ausgewertet. Diese Integrale findet man in Pearson & Hartley (1976) tabelliert, und zwar für die Werte
$zs = 0.0\ (0.05)\ 7.50$, $p = 2(1)20$. Nach Hartley (1942) läßt sich die Verteilungsfunktion der Spannweite in (4.51) berechnen durch

(4.55) $\forall p \geq 2 : \forall z \geq 0 :$

$$G_p(z) = [\Phi(z/2) - \Phi(-z/2)]^p + 2p \int_{z/2}^{\infty} \varphi(x)\, [\Phi(x) - \Phi(x-z)]^{p-1}\, dx .$$

Der erste Summand wurde mit Hilfe der Formeln von Kerridge & Cook (1976) für Φ berechnet. Entsprechendes gilt für den Integranden im zweiten Summanden. Das Integral wurde anschließend mit der Simpson Formel ($h = 0.01$) ausgewertet. Die obere Integrationsgrenze wurde auf 7.5 festgelegt. Zur Berechnung von (4.50) wurde $G_p(z)$ für $z = 0(0.01)7.5$, $p = 2,\ldots,20$ tabelliert. Für $z > 7.5$ ist $G_p(z)$ für $p \leq 20$ annähernd gleich 1 (vgl. Pearson, Hartley (1976)).

In (4.50) läßt sich nun $I(zs) := \prod_{i=1}^{l} G_{p_i}(zs)$ durch Multiplikation der entsprechenden Tabellen für $G_{p_i}(zs)$ für $zs = 0.00\ (0.01)\ 7.50$ tabellieren.

Mit der Substitution $z = (\nu s)^2$ und $I(zs) \approx 1$ für $zs \geq 7.5$ erhält man

$$G_{(p_1,\ldots,p_l;\nu)}(z) \approx \int_0^{7.5/z} I(zs)\, f_\nu(s)\, ds + P(\chi_\nu^2 > \nu\, (\tfrac{7.5}{z})^2) .$$

Der erste Summand wurde mit der Simpson–Formel ($h = 0.01$) ausgewertet, während $P(\chi_\nu^2 > \nu\, (\frac{7.5}{z})^2)$ ohne Probleme mit Hilfe der Formeln von Lackritz (1984) berechnet wurde. Zur Berechnung der kritischen Werte der studentisierten Maximum–Range-Verteilung, die mit $q_{(p_1,\ldots,p_l;\nu),\alpha}$ bezeichnet werden, wurde der Aitken–Neville-Algorithmus verwendet. Da die Verteilungsfunktion der Spannweite bei der Normalverteilung log–konkav in k ist (vgl. Satz 2.1), erfüllen auch die kritischen Werte der studentisierten Maximum–Range–Verteilung die Monotonieeigenschaft (4.40).

Zu beachten ist allerdings, daß sich für unterschiedliche ν unterschiedliche Ordnungen der kritischen Werte ergeben können. So ist zum Beispiel für $\nu = 1$

$q_{(7,3,2;1),0.05} = 46.18 < q_{(5,4,4;1),0.05} = 46.521$, aber für $\nu = \infty$

$q_{(7,3,2;\infty),0.05} = 4.269 > q_{(5,4,4;\infty),0.05} = 4.247.$

Für die praktische Durchführung der Testprozedur im ANOVA–Modell wurde ein FORTRAN–Programm entwickelt, welches Probleme mit bis zu 20 Mittelwerten löst und vom Autor erhältlich ist. Die Rechenzeit auf einem PC liegt im Minutenbereich. Die Durchführung ohne Rechnerhilfe ist für große k sicher problematisch. Ein Beispiel mit 13 Mittelwerten zeigt jedoch eindrucksvoll, wie stark sich die Anzahl der tatsächlich zu prüfenden Hypothesen unter Ausnutzung aller in Abschnitt 4.2 erzielten Resultate verringern läßt. Die folgende vorbereitende Bemerkung dient der Vereinfachung der Schreibweise.

Bemerkung 4.4. Die Struktur der Elemente $(P_1,\ldots,P_1) \in \mathfrak{Q}_2^{**}(u,v)$ erlaubt es, die Mengen P_i durch das Minimum und Maximum der Elemente von P_i zu beschreiben und damit die gesamte Partition $(P_1,\ldots,P_1)$ eindeutig festzulegen. Wir verzichten auf eine formale Definition und geben dafür ein sich selbst erklärendes Beispiel. Sei $k = 13$, $u = 5$, $v = 9$ und $P_1 = \{1,2,3\}$, $P_2 = \{4,6,7\}$, $P_3 = \{5,9\}$, $P_4 = \{8,10,11,12,13\}$. Dann schreiben wir für $(P_1,\ldots,P_4)$: $(1,3)(4,7)(5,9)^*(8,13)$. Dabei wird die Menge, die u und v enthält, durch einen * markiert.

Beispiel 4.2. Wir wählen das schon in der Einleitung betrachtete Beispiel von Keuls (1952) mit 13 Mittelwerten:

i	1	2	3	4	5	6	7	8	9	10	11	12	13
$\bar{x}_{i.}$	97.7	100.7	111.3	120.7	124.3	128.7	129.0	131.0	132.0	141.7	150.7	152.7	176.0

Die Mittelwerte wurden aus 3 Beobachtungen gebildet, die Varianz mit $s^2 = 124.29$ geschätzt ($\nu = 24$). Die t–Prüfgrößen sowie die Testergebnisse des Tukey–Tests und des Newman–Keuls Tests sind der Tabelle 4.2 zu entnehmen. Tabelle 4.3 gibt die 26 größten der insgesamt 100 benötigten kritischen Werte für $\alpha = 0.05$ wieder.

Tabelle 4.2:

j \ i	1	2	3	4	5	6	7	8	9	10	11	12
2	0.466											
3	2.112	1.647										
4	3.573	3.107	1.460									
5	4.133	3.667	2.020	0.560								
6	4.816	4.350	2.703	1.243	0.684							
7	4.863	4.397	3.750	1.290	0.730	0.047						
8	5.174	4.707	3.061	1.600	1.041	0.357	0.311					
9	5.329	4.863	3.216	1.756	1.196	0.513	0.466	0.155				
10	6.836	6.370	4.723	3.263	2.703	2.020	1.973	1.662	1.507			
11	8.234	7.768	6.121	4.661	4.102	3.418	3.371	3.061	2.905	1.398		
12	8.545	8.079	8.432	4.972	4.412	3.729	3.682	3.371	3.216	1.706	0.311	
13	12.165	11.699	10.052	8.591	8.032	7.349	7.302	6.991	6.836	5.329	3.931	3.620

t–Statistiken für das Keuls–Beispiel

_____ Tukey lehnt ab.

- - - - - Newman–Keuls lehnt zusätzlich ab (strittige Hypothesen)

Tabelle 4.3:

$p_1,\ldots,p_l$	$q_{(p_1,\ldots,p_l;24),0.05}$
13	5.179
12	5.099
11,2	5.029
11	5.012
10,3	4.969
10,2	4.935
9,4	4.923
10	4.915
8,5	4.891
7,6	4.875
9,3	4.872
9,2,2	4.855
9,2	4.832

$p_1,\ldots,p_l$	$q_{(p_1,\ldots,p_l;24),0.05}$
8,4	4.825
9	4.807
7,5	4.796
8,3,2	4.791
6,6	4.786
8,3	4.765
7,4,2	4.746
8,2,2	4.746
7,3,3	4.729
6,5,2	4.723
7.4	4.718
8,2	4.715
6,5	4.694

Nach Durchführung des Tukey–Tests und des Newman–Keuls Test bleiben 11 Hypothesen als strittig übrig. Indem wir den Newman–Keuls Test (die obere Schranke) wie beschrieben in einfacher Weise verbessern, können wir diese Anzahl wie folgt verringern.

Wählt man $(P_1,P_2,P_3) = (1,2)(3,10)(11,13)$, so ist $t_{(P_1,P_2,P_3)} = 4.723 < q_{(8,3,2;24),0.05} = 4,791$ und damit $\bar{\varphi}_{11,13} = \bar{\varphi}_{12,13} = \bar{\varphi}_{3,10} = 0$.

Für $(P_1,P_2,P_3,P_4) = (1)(2,8)(9,12)(13)$ erhält man $t_{(P_1,\ldots,P_4)} = 4.707 < q_{(7,4;24),0.05} = 4,718$, also $\bar{\varphi}_{2,6} = \bar{\varphi}_{2,7} = \bar{\varphi}_{2,8} = 0$.

Damit bleiben noch 5 strittige Hypothesen zu prüfen, die in der Reihenfolge (1,8) , (4,12) , (2,9) , (1,7) , (1,6) abgearbeitet werden. Auf das Problem der Bestimmung der Mengen $Q_2^{**}(u,v)$ soll hier nicht weiter eingegangen werden. Zunächst erhält man $Q_2^{**}(1,8) = Q_2^{**}(4,12) = \emptyset$, und damit $\bar{\varphi}_{1,8} = \bar{\varphi}_{4,12} = 1$. $Q_2^{**}(2,9)$ hat lediglich ein Element, nämlich $(1)(2,9)^*(3,13)$. Wegen $\bar{\varphi}_{3,13} = 1$ folgt sofort

$\varphi_{(1)(2,9)^*(3,13)} = 1$, also $\bar{\varphi}_{2,9} = 1$. $Q_2^{**}(1,7)$ besteht aus den Partitionen

$\varphi_{(1)(2,9)^*(3,13)} = 1$, also $\bar{\varphi}_{2,9} = 1$. $Q_2^{**}(1,7)$ besteht aus den Partitionen $(1,7)^*(2,13)$, $(1,7)^*(3,13)$, $(1,7)^*(2,12)(13)$, $(1,7)^*(4,13)$, $(1,7)^*(5,13)$, $(1,7)^*(6,13)$ und $(1,7)^*(3,12)(13)$. Da diese Partitionen alle Elementarhypothesen enthalten, die schon abgelehnt worden sind, folgt sofort $\bar{\varphi}_{1,7} = 1$. Die gleiche

Argumentation gilt für $Q_2^{**}(1,6)$, die aus den Partitionen $(1,6)^*(3,13)$, $(1,6)^*(4,13)$, $(1,6)^*(3,12)(13)$, $(1,6)^*(4,13)$, $(1,6)^*(5,13)$, $(1,6)^*(7,13)$, $(1,6)^*(3,12)(13)$, $(1,6)^*(2,11)(12,13)$, $(1,6)^*(2,11)(12)(13)$ und $(1,6)^*(4,12)(13)$ besteht. Damit ist auch $\bar{\varphi}_{1,6} = 1$.

Insgesamt werden im Verhältnis zum Tukey–Test 5 weitere Paarhypothesen abgelehnt. Der kleinste t–Wert der zusätzlich abgelehnten Hypothesen ist $t_{1,6} = 4.816$. Aus $G_{13}(4.816) = 0.914$ ergibt sich, daß man den Tukey–Test zum Niveau $\alpha = 0.086$ hätte durchführen müssen, um die beim Abschlußtest zusätzlich abgelehnten Hypothesen auch mit einem Tukey–Test ablehnen zu können. Durch den Abschluß–Test wird also in diesem Beispiel ein Niveaugewinn von ca. 70 % erzielt. □

5. Abschlußbemerkungen

Eine wichtige Voraussetzung der in dieser Arbeit betrachteten Lokationsmodelle war, daß die X_i unter der Globalhypothese die gleiche Verteilung besitzen. Das ist in der statistischen Praxis leider nicht immer der Fall. Hier ist insbesondere das Auftreten unterschiedlicher Varianzen zu nennen. Diese Situation entsteht auch, wenn etwa ein ANOVA–Modell mit ungleichen Stichprobenumfängen vorliegt. Für diesen Fall gibt es in der Literatur eine Reihe von Vorschlägen für simultane Testprozeduren (vgl. etwa Tukey (1953), Scheffé (1953), Kramer (1956), Dunn (1974), Hochberg (1974), Stoline (1978), Gabriel (1978)). Simulationsstudien von Dunnett (1980) und Stoline (1981) haben ergeben, daß die lange Zeit umstrittene Tukey–Kramer Methode unter den Verfahren, die das multiple Niveau α einhalten, am besten abschneidet. Hayter (1984) konnte zeigen, daß die Tukey–Kramer Methode das multiple Niveau α einhält. Da die Ordnungseigenschaften der entsprechenden t–Statistiken bei diesem Ansatz im allgemeinen verlorengehen, wird die Durchführung von sequentiell verwerfenden Varianten der Tukey–Kramer Methode erheblich schwieriger. Insbesondere erscheint eine dem Kapitel 4 entsprechende Variante des Abschlußtests für größere k kaum berechenbar zu sein. Da die Tukey–Kramer Methode auf einer Ungleichung beruht, wird das multiple Niveau beim Vorliegen ungleicher Stichprobenumfänge in keinem Fall ausgeschöpft. Für $k = 3$ konnten Spurrier, Isham (1985) die exakten kritschen Werte für die entsprechenden Statistiken berechnen, so daß in diesem Fall ein Abschlußtest das multiple Niveau ausschöpft.

Einen ganz anderen Ansatz liefert die Verwendung von χ^2–Tests oder F–Tests. Obwohl die Berechnung der entsprechenden Statistiken erheblich aufwendiger ist (dabei ist an die von Cox, Spjotvoll (1981) in einem anderen Zusammenhang verwendeten Statistiken für die Partitionshypothesen gedacht), gibt es auch für diese Statistiken eine Reihe von Ordnungseigenschaften (einige davon findet man schon in Gabriel (1963)),

die man bei der Durchführung eines Abschlußtests ausnutzen kann. Ein Vorteil beim Verwenden von F–Tests ist, daß man nur k–1 verschiedene kritische Werte benötigt. Erste Versuche zeigen, daß auch hier bei gleichen Stichprobenumfängen ein Abschlußtest mit Rechnerhilfe bis zu etwa 20 Mittelwerten durchführbar ist. Bei ungleichen Stichprobenumfängen hingegen ist mit noch größeren Schierigkeiten als bei den Spannweitentests zu rechnen. Dabei wäre ein Einsatz von F–Tests gerade hier lukrativ, da im Gegensatz zur Tukey–Kramer Methode immer die exakten Verteilungen bzw. kritischen Werte benutzt werden können und somit das multiple Niveau voll ausgeschöpft wird. Ein Nachteil von multiplen F–Tests (oder χ^2–Tests) ist, daß sie im allgemeinen nicht lokal konsonant sind (vgl. Beispiel 1.1). Das gilt auch, wie sich leicht zeigen läßt, für die häufig angeführte F–Variante des abgeschlossenen Newman–Keuls Tests für 3 Mittelwerte (wähle etwa $\alpha = 0.001$ und $\nu = 24$).

Ein bis heute speziell für die Paarvergleiche weitgehend ungelöstes Problem beim Verwenden von sequentiell verwerfenden Testverfahren ist die Kontrolle der sogenannten Fehler III. Art, die immer dann auftreten können, wenn man sich bei Ablehnung einer Hypotnese $H_0^{ij} : \mu_i = \mu_j$ nicht nur für $H_0^{ij} : \mu_i \neq \mu_j$, sondern zusätzlich für $\mu_i < \mu_j$ oder $\mu_i > \mu_j$ entscheidet. Ein Fehler III. Art tritt genau dann auf, wenn man sich für $\mu_i < \mu_j$ entscheidet, $\mu_i > \mu_j$ jedoch richtig ist oder umgekehrt. Obwohl derartige Entscheidungen bei der in dieser Arbeit gewählten Hypothesenformulierung nicht vorgesehen sind, wird der Anwender darauf kaum Rücksicht nehmen. So stellt sich die Frage, ob die hier vorgestellten sequentiell verwerfenden Testprozeduren sowohl die Fehler I. Art als auch die Fehler III. Art durch das multiple Niveau α kontrollieren. Shaffer (1980) konnte für eine andere Hypothesenfamilie zeigen, daß die Antwort von den zugrundeliegenden Verteilungen abhängt (vgl. auch Holm (1979b)). Das gilt wohl auch für die Paarvergleiche. So läßt sich für die Normalverteilung bei 3 Mittelwerten zeigen, daß bei Verwendung des Abschlußtests auch die Fehler III. Art durch das multiple Niveau α kontrolliert werden, während dies bei der Cauchy–Verteilung nicht

der Fall ist (vgl. auch Finner (1988)). Von einer allgemeinen Beantwortung dieser Fragestellung ist man jedoch noch weit entfernt.

Ein weiteres prinzipielles Problem ist der Vergleich der Güte von multiplen Testverfahren. Hier stellt sich insbesondere die Frage, wie man die mehrdimensionalen Gütefunktionen verschiedener multipler Tests miteinander vergleichen soll und welche Funktionale der Gütefunktionen gegebenenfalls als Vergleichskriterien heranzuziehen sind. Selbst wenn man ein geeignetes Funktional gewählt hat, ist ein theoretischer Nachweis, ob oder wann die eine oder andere Testprozedur besser ist, kaum zu führen. Einige theoretische Ansätze zu dieser Problemstellung findet man in der Arbeit von Spjotvoll (1972). Simulationsstudien, wie etwa von Carmer, Swanson (1973), Ramsey (1978), Einot, Gabriel (1975) oder auch Shaffer (1981) haben den Nachteil, daß sie nur für wenige ausgewählte Parameterkonfigurationen durchgeführt werden können und so leicht zu falschen Schlüssen führen können. Um auf diesem Gebiet befriedigende Ergebnisse zu erzielen, ist wohl noch einige Arbeit notwendig. □

Literaturverzeichnis:

ABRAMOWITZ, M., STEGUN, I.A. (1968). Handbook of Mathematical Functions. Dover Publications, Inc., New York.

BAUER, H. (1978). Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie. 2. neubearb. Aufl., Berlin, New York: de Gruyter.

BEGUN, J.M., GABRIEL, K.R. (1981). Closure of the Newman–Keuls multiple comparisons procedure. J. Am. Statist. Assoc. 76, 241–245.

BILLINGSLEY, P. (1968). Convergence of Probability Measures. Wiley, New York.

BILLINGSLEY, P. (1979). Probability and Measure. Wiley, New York.

BURR, I.W. (1970). Calculations of exact sampling distribution of ranges from a discrete population. Ann. Math. Statist. 26, 530–532. Correction 38, 280.

CARMER, S.G., SWANSON, M.R. (1973). An evaluation of ten pairwise multiple comparison procedures by Monte Carlo methods. J. Am. Statist. Assoc. 68, 66–74.

COX, D.R., SPJOTVOLL, E. (1982). On partitioning means into groups. Scand. J. Statist. 9, 147–152.

DAVID, H.A. (1970). Order Statistics. John Wiley & Sons, Inc., New York, London, Sydney, Toronto.

DUNCAN, D.B. (1955). Multiple range and multiple F–tests. Biometrics 11, 1–42.

DUNN, O.J. (1974). On multiple tests and confidence intervals. Comm. Statist. 3, 101–103.

DUNNETT, C.W. (1980). Pairwise multiple comparisons in the homogeneous variance, unequal sample size case. J. Am. Statist. Assoc. 75, 789–795.

EINOT, I., GABRIEL, K.R. (1975). A study of the powers of several methods of multiple comparisons. J. Am. Statist. Assoc. 70 , 574–583.

FINNER, H. (1988). Multiple Tests und Fehler III. Art. Erscheint in: Bauer, P. et al. (Hrsg.): Multiple Hypothesenprüfungen. Springer: Berlin, Heidelberg, New York.

FISHER, R.A. (1935). The Design of Experiments. Oliver & Boyd, Ltd., Edinburgh and London.

GABRIEL, K.R. (1963). A procedure for testing the homogeneity of all sets of means in analysis of variance. Biometrics 20, 459–477.

GABRIEL, K.R. (1969). Simultaneous test procedures – some theory of multiple comparisons. Ann. Math. Statist. 40, 224–250.

GABRIEL, K.R. (1978). A simple method of multiple comparisons of means. J. Am. Statist. Assoc. 73, 724–729.

GALTON, F. (1902). The most suitable proportion between the values of first and second prizes. Biometrika 1, 385–390.

GUMBEL, E.J. (1958). Statistics of Extremes. Columbia University Press. New York, Basel.

HARTER, H.L. (1960). Tables of range and studentized range. Ann. Math. Statist. 31, 1122–1147.

HARTLEY, H.O. (1942). The range in random samples. Biometrika 32, 334–348.

HARTLEY, H.O. (1955). Some recent developments in analysis of variance. Comm. Pure Appl. Math. 8, 47–72.

HAYTER, A.J. (1984). A proof of the conjecture that the Tukey–Kramer multiple comparisons procedure is conservative. Ann. Statist. 12, 61–75.

HAYTER, A.J. (1986). The maximum familywise error rate of Fisher's least significance difference test. J. Am. Statist. Assoc. 81, 1000–1004.

HOCHBERG, Y. (1974). Some conservative generalizations of the T–method in simultaneous inference. J. Mult. Anal. 4, 224–234.

HOLM, S. (1977). Sequentially rejective multiple test procedures. Statistical Research Report 1977–1, University of Umea, Sweden.

HOLM, S. (1979a). A simple sequentially rejective multiple test procedure. Scand. J. Statist. 6, 65–70.

HOLM, S. (1979b). A stagewise directional test based on t statistics. Research Report 1979–3, Institute of Mathematics, Chalmers University of Technology, Gothenburg.

HOMMEL, G. (1986). Multiple test procedures for arbitrary dependence structures. Metrika 33, 321–336.

HOMMEL, G. (1988). A stagewise rejective multiple test procedure based on a modified Bonferroni test. Erscheint in Biometrika 75.

KERRIDGE, D.F., COOK, G.W. (1976). Yet another series for the normal integral. Biometrika 63, 401–403.

KEULS, M. (1952). The use of the studentized range in connection with an analysis of variance. Euphytica 1, 112–122.

KIMBALL, A.W. (1951). On dependent tests of significance in the analysis of variance. Ann. Math. Statist. 22, 600–602.

KRAMER, C.Y. (1956). Extension of multiple range tests to group means with unequal numbers of replications. Biometrics 12, 307–310.

LACKRITZ, J.R. (1984). Exact p–values for F and t–tests. Am. Statistician 38, 312–314.

LEHMANN, E.L. (1957a). A theory of some multiple decision problems I. Ann. Math. Statist. 28, 1–25.

LEHMANN, E.L. (1957b). A theory of some multiple decision problems II. Ann. Math. Statist. 28, 547–572.

LEHMANN, E.L. (1966). Some concepts of dependence. Ann. Math. Statist. 37, 1137–1153.

LEHMANN, E.L., SHAFFER, J.P. (1977). On a fundamental theorem in multiple comparisons. J. Am. Statist. Assoc. 72, 576–578.

LEHMANN, E.L., SHAFFER, J.P. (1979). Optimum significance levels for multistage comparison procedures. Ann. Stat. 7, 27–45.

LEHMANN, E.L. (1986). Testing Statistical Hypotheses. Second Edition. Wiley, New York.

MARCUS, R., PERITZ, E., GABRIEL, K.R. (1976). On closed testing procedures with special reference to ordered analysis of variance. Biometrika 63, 655–660.

MILLER, R.G. (1966). Simultaneous Statistical Inference. New York: Mc Graw–Hill.

MILLER, R.G. (1981). Simultaneous Statistical Inference. Second Edition. Springer–Verlag, New York, Heidelberg, Berlin.

NAIK, U.D. (1975). Some selection rules for comparing p procedures with a standard. Comm. Statist. 4, 519–535.

NEWMAN, D. (1939). The distribution of range in samples from a normal population, expressed in terms of an independent estimate of standard deviation. Biometrika 31, 20–30.

O'NEILL, R., WETHERILL, G.B. (1971). The present state of multiple comparison methods. With discussion. J. R. Statist. Soc. B 33, 218–250.

PATEL, J.K., KAPADIA, C.H., OWEN, D.B. (1976). Handbook of Statistical Distributions. Marcel Dekker, Inc., New York, Basel.

PEARSON, E.S. (1926). A further note on the distribution of range in samples taken from a normal distribution. Biometrika 18, 173–194.

PEARSON, E.S., HARTLEY, H.O. (1976). Biometrika Tables for Statisticians. Biometrika Trust.

PEARSON, K. (1902). Note on Francis Galton's problem. Biometrika 1, 484–487.

PERITZ, E. (1970). A note on multiple comparisions. Unpublished Paper. Hebrew University.

RAMSEY, P.H. (1978). Power differences between pairwise multiple comparisons. J. Am. Statist. Assoc. 73, 479–487.

ROY, S.N. (1953). On a heuristic method of test construction and its use in a multivariate analysis. Ann. Math. Statist. 24, 220–238.

ROY, S.N., BOSE, R.C. (1953). Simultaneous confidence interval estimation. Ann. Math. Statist. 24, 513–536.

ROYEN, T. (1987). Eine verschärfte Holm–Prozedur zum Vergleich aller Mittelwertpaare. EDV in Med. u. Biologie 18, 45–49.

RYAN, T.A. (1959). Multiple comparisons in psychological research. Psychological Bulletin 56, 26–47.

RYAN, T.A. (1960). Significance tests for multiple comparison of proportions, variances, and other statistics. Psychological Bulletin 57, 318–328.

SAS INSTITUTE INC. (1985). SAS/STAT™ Guide for Personal Computers, Version 6. Edition. Cary, NC: SAS Institute Inc..

SCHEFFE, H. (1953). A method for judging all contrasts in the analysis of variance. Biometrika 40, 87–104.

SHAFFER, J.P. (1980). Control of directional errors with stagewise test procedures. Ann. Statist. 8, 1342–1347.

SHAFFER, J.P. (1981). Complexity: An interpretability criterion for multiple comparisons. J. Am. Statist. Assoc. 76, 395–401.

SHAFFER, J.P. (1984). Issues arising in multiple comparisons among population. Proceedings of the Seventh Conference on Probability Theory, 353–362. Ed. M. Josifescu, Bucharest, Romania: Editoria Acadimiei Republicii Socialiste Romania.

SHAFFER, J.P. (1986). Modified sequentially rejective multiple test procedures. J. Am. Statist. Assoc. 81, 826–831.

SIMES, R.J. (1986). An improved Bonferroni procedure for multiple tests of significance. Biometrika 73, 751–754.

SONNEMANN, E. (1982). Allgemeine Lösungen multipler Testprobleme. EDV in Med. und Biologie 13, 120–128.

SONNEMANN, E., FINNER, H. (1988). Vollständigkeitssätze für multiple Testprobleme. Erscheint in: Bauer, P. et al. (Hrsg.): Multiple Hypothesenprüfungen. Springer: Berlin, Heidelberg, New York.

SPJOTVOLL, E. (1972). On the optimality of some multiple comparison procedures. Ann. Math. Statist. 43, 398–411.

SPURRIER, J.D., ISHAM, ST.P. (1985). Exact simultaneous confidence intervals for pairwise comparisons of three normal means. J. Am. Statist. Assoc. 80, 438–442.

STOLINE, M.R. (1978). Tables of the studentized augmented range and applications to multiple comparison. J. Am. Statist. Assoc. 73, 656–660.

STOLINE, M.R. (1981). The status of multiple comparisons: Simultaneous estimation of all pairwise comparisons in one–way ANOVA designs. Am. Statistician 35, 134–141.

STUDENT (1927). Errors in routine analysis. Biometrika 19, 151–164.

TIPPETT, L.H.C. (1925). On the extreme individuals and the range of samples taken from a normal population. Biometrika 17, 364–387.

TUKEY, J.W. (1949). Comparing individual means in the analysis of variance. Biometrics 5, 99–114.

TUKEY, J.W. (1952). Allowances for various types of error rates. Unpublished IMS address, Virginia Polytechnic Institute, Blacksburg, Va.

TUKEY, J.W. (1953). The problem of multiple comparisons. Unpublished manuscript.

WELSCH, R.E. (1977). Stepwise multiple comparison procedures. J. Am. Statist. Assoc. 7, 566–575.

Symbole und Bezeichnungen

$\mathbb{N}$	Menge der natürlichen Zahlen
$\mathbb{Z}$	Menge der ganzen Zahlen
$\mathbb{R}$	Menge der reellen Zahlen
2^A	Potenzmenge von A
$\mathfrak{X}$	Stichprobenraum
$\mathcal{B}$	σ–Algebra über $\mathfrak{X}$
θ	Parameterraum
$A \cup B$ bzw. $A \cap B$	Vereinigung bzw. Durchschnitt von Mengen A und B
$A + B$	disjunkte Vereinigung von A,B
ΣA_i	disjunkte Vereinigung von Mengen A_i
$A \subset B$	A Teilmenge von B und $A \neq B$
$A \subseteq B$	$A \subset B$ oder $A = B$
$\|A\|$	Kardinalzahl von A
$\supset, \supseteq$	analog zu $\subset, \subseteq$
X, X_i, etc.	Zufallsvariable
$X \sim \ldots$	X verteilt wie ...
$N(\mu, \sigma^2)$	Normalverteilung mit Erwartungswert μ und Varianz σ^2
χ^2_ν	Chi–Quadratverteilung mit ν Freiheitsgraden
R_k	Spannweite von k stochastisch unabhängig identisch verteilten Zufallsvariablen
G_k	Verteilungsfunktion von R_k
$G_{(p_1,\ldots,p_l)}$	S. 58
$c_\alpha(p), c_\alpha(p_1,\ldots,p_l)$, etc.	kritische Werte
$\chi^2_{\nu,\alpha}$	α–Fraktil der χ^2–Verteilung mit ν Freiheitsgraden

u_α	α–Fraktil der $N(0,1)$–Verteilung
$q_{k,\nu,\alpha}$	α–Fraktil der studentisierten Spannweite
$q_{(p_1,\ldots,p_l;\nu,\alpha)}$	α–Fraktil der verallgemeinerten studentisierten Spannweite
$H_0^i, H_0^{ij}, H_0^{(P_1,\ldots,P_l)}, H_0^P$	Nullhypothesen
H_1^i, H_1^{ij}	Alternativhypothesen
H_0^P	Homogenitätshypothesen
$H_0^\pi, H_0^{(P_1,\ldots,P_l)}$	Partitionshypothesen
$I(\vartheta)$	S. 5
$\mathcal{H}$	Menge von Hypothesen (Paarhypothesen)
$\mathcal{H}^*$	Menge der Homogenitätshypothesen
$\overline{\mathcal{H}}$	$\cap$–abg. Hypothesenfamilie
Π_k, N_k	S. 36
$L(p_1,\ldots,p_l)$	S. 39
$\Pi, (P_1,\ldots,P_l)$	Partitionen von $I = \{1,\ldots,k\}$
$Q(u,v), Q^*(u,v)$	S. 62, S. 63
$Q_1^{**}(u,v), Q_2^{**}(u,v)$	S. 72
min P, max P	S. 63
$d(P), d(P_1,\ldots,P_l)$	Durchmesser von P bzw. $(P_1,\ldots,P_l)$, S. 64
P_r^*, M_r^*, P_s^*	S. 63, S. 65, S. 70
$\varphi, \varphi_i, \varphi_P, \varphi_{(P_1,\ldots,P_l)}$, etc.	Tests
Φ_α	S. 44
$\overline{\varphi}$	Abschlußtest
$\overline{\Psi}$	S. 57
$\xrightarrow{w}$	konvergiert schwach gegen
$\exists, \forall$	es existiert, für alle
ANOVA	Analysis of Variance
$\cap$–abg.	durchschnittsabgeschlossen